## (12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification[7]: G01N 33/48

(21) International Application Number: PCT/US02/09051

(22) International Filing Date: 12 March 2002 (12.03.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/275,144   12 March 2001 (12.03.2001)   US

(71) Applicant (for all designated States except US): BOARD OF REGENTS, THE UNIVERSITY OF TEXAS SYSTEM [US/US]; 201 W. 7th Street, Austin, TX 78701 (US).

(72) Inventors; and
(75) Inventors/Applicants (for US only): FOX, Robert, O.

[US/US]; 8 Quintana Drive, Galveston, TX 77554 (US). YANG, Huan-wang [CN/US]; 17 Kearny Ave. 5B, Edison, NJ 08817 (US).

(74) Agent: ACOSTA, Melissa, W.; Fulbright & Jaworski L.L.P., 1301 McKinney, Suite 5100, Houston, TX 77010-3095 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),

(54) Title: COMPUTER-BASED STRATEGY FOR PEPTIDE AND PROTEIN CONFORMATIONAL ENSEMBLE ENUMERATION AND LIGAND AFFINITY ANALYSIS



The peptide
RCCHPQCGMVEECR
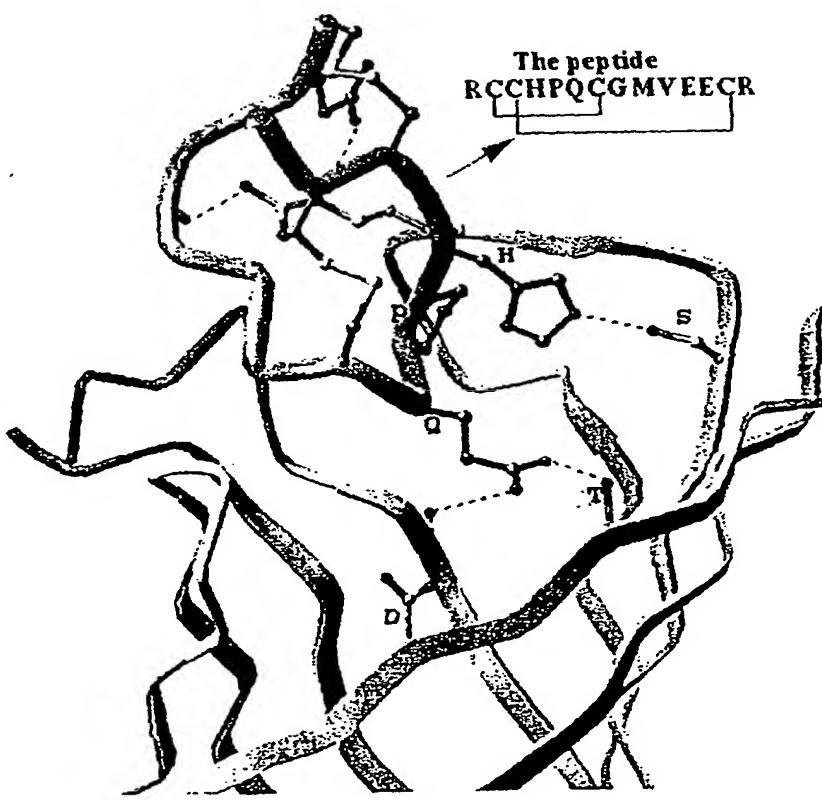
(57) Abstract: The present invention provides a method to generate and analyse ensembles of peptide and protein conformers and predict the affinity of a given conformation of the peptide or protein for a target protein.

WO 02/073193 A1

European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Published:**

   —   *with international search report*

—  *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

# COMPUTER-BASED STRATEGY FOR PEPTIDE AND PROTEIN CONFORMATIONAL ENSEMBLE ENUMERATION AND LIGAND AFFINITY ANALYSIS

## BACKGROUND OF THE INVENTION

[0001] This application claims priority to U.S. Provisional Application No. 60/275,144, which was filed on March 12, 2001.

[0002] The U.S. Government may have certain rights in the invention by virtue of a grant from DARPA.

## I.      Field of the Invention

[0003] The invention generally relates to the field of structural biology. It concerns a method of modeling the structure of a peptide and stabilizing the structure of that peptide by the insertion of an amino acid not naturally found in that position in the peptide. It also concerns a method for assessing the binding affinity of a peptide to a template molecule and a method for determining the rate of loop closure in a peptide via a disulfide bond.

## II.      Related Art

[0004] The protein modeling approach of the present invention provides an efficient method of predicting where to insert cysteines or other amino acids in a peptide in order to stabilize the peptide.

[0005] The three-dimensional structure of proteins has been determined in a number of ways. One of the most well known way of determining protein structure involves the use of the technique of x-ray crystallography. Using this technique, it is possible to elucidate the three-dimensional structure with good precision. Additionally, protein structure may be determined through the use of the techniques of neutron diffraction, or by nuclear magnetic resonance (NMR).

[0006] The three-dimensional structure of many proteins may be characterized as having internal surfaces (directed away from the aqueous environment in which the protein is normally found) and external surfaces (which are exposed to the aqueous environment). Through the study of many natural proteins, researchers have discovered that hydrophobic residues (such as tryptophan, phenylalanine, leucine, isoleucine, valine, or methionine) are most frequently found on the internal surface of protein molecules. In contrast, hydrophilic residues (such as aspartate, asparagine, glutamate, glutamine, lysine, arginine, serine, and threonine) are most frequently found on the external protein surfaces. The amino acids

alanine, cysteine, glycine, histidine, proline, serine, tyrosine, and threonine are encountered with more nearly equal frequency on both the internal and external protein surfaces.

[0007] The biological properties of proteins depend directly on the protein's three-dimensional (3D) conformation. The conformation determines the activity of enzymes, the capacity and specificity of binding proteins, and the structural attributes of receptor molecules. Each protein has an astronomical number of possible conformations (about 1016 for a small protein of 100 residues, and there has been no reliable method for picking the one conformation that predominates in aqueous solution. A second difficulty is that there are no accurate and reliable force laws for the interaction of one part of a protein with another part and with water. These and other factors have contributed to the enormous complexity of determining the most probable relative location of each residue in a known protein sequence.

[0008] The protein folding problem, the problem of determining a protein's three-dimensional tertiary structure from the amino acid sequence, was first formulated more than half a century ago. Early observations and later experiments have lead to the contemporary view that protein conformation is determined solely by the amino acid sequence and that there exists a unique native conformation in which residues distant in sequence but proximate in space engender a close-packed core enriched in hydrophobic residues. As a result of the revolution in molecular biology, the number of known protein sequences is about 50 times greater than the number of known three-dimensional protein structures. This disparity hinders progress in many areas of biochemistry because a protein sequence has little meaning outside the context of the three-dimensional structure.

[0009] The protein modeling approach of the present invention provides an efficient method of predicting where to insert cysteines or other amino acids in a peptide in order to stabilize the peptide. The present invention provides an ensemble-based all-atom method, mini-protein modeling program (MPMOD), to stabilize a protein to provide higher affinity binding.

## SUMMARY OF THE INVENTION

[0010] Therefore, it is an objective of the present invention to provide a method to generate and analyze ensembles of peptide and protein conformers and predict the affinity of a given conformation of the peptide or protein for a target protein.

[0011] An embodiment of the present invention is a computer-assisted method for use in modifying a protein comprising the steps of: inputting a peptide sequence and parameters for analysis into a computer-assisted modeling program; randomly generating $\phi$, $\psi$, $\omega$;

2

generating a peptide backbone; performing a van der Waals check of the backbone; calculating a solvent accessible surface based energy of all conformers; modeling the disulfide(s); performing a van der Waals check after each rotamer is added; and calculating a solvent accessible surface based energy of all conformers. In a further embodiment, the protein may comprise a peptide. In yet a further embodiment, the method may comprise performing a binding test for each conformer with a template molecule. In a further embodiment, the method may comprise calculating the rate of disulfide bond loop closure.

[0012] Another embodiment of the invention is a method of protein miniaturization comprising modeling a protein to have the necessary active site conformation using the method above while reducing the total number of amino acids in the protein.

[0013] Yet another embodiment is a method of increasing binding affinity between a protein and a template molecule by decreasing the conformational entropy loss upon binding by the protein comprising the constraint of at least one loop of an unstable region of the protein in conformational space using the method above.

[0014] Another embodiment of the present invention is a computer-assisted method for use in modifying a protein comprising the steps of: inputting a peptide sequence into a computer-assisted modeling program; inputting parameters for analysis into a computer-assisted modeling program; generating $\phi$, $\psi$, $\omega$ angles randomly in allowed region of Ramachandran maps; assigning angles to each residue of the backbone; generating backbone atoms for N, CA, C; generating the rest of the backbone atoms; performing van der Waals check for each atom; modeling disulfide bonds and recording the disulfide coordinate pairs; adding rotamers to residues; performing van der Waals check for each rotamer; performing binding test with a template protein; and calculating solvent accessible surface based energy for each conformer. In a further embodiment the protein may comprise a peptide.

[0015] Another embodiment of the present invention is a computer-assisted method for use in modifying a protein comprising the steps of: inputting a peptide sequence into a computer-assisted modeling program; inputting parameters for analysis into a computer-assisted modeling program; generating $\phi$, $\psi$, $\omega$ angles randomly in allowed region of Ramachandran maps; assigning angles to each residue of the backbone; generating backbone atoms N, CA, C; generating the rest of the backbone atoms; performing van der Waals check with all other atom after each atom is added; adding rotamers to residues; checking distance pairs between atoms; modeling disulfide bonds and recording the disulfide

3

coordinate pairs; performing van der Waals check for the disulfide bonds with the complete conformer; recording number of conformers that are able to form disulfide bonds; and calculating solvent accessible surface based energy for each conformer.

[0016] Yet another embodiment of the present invention is a computer-assisted method for use in modifying a protein comprising the steps of: inputting a peptide sequence and parameters for analysis into a computer-assisted modeling program; searching of conformational space in the allowed regions of the Ramachandran plots; minimizing the N and C termini of the conformer to be the same as the high resolution structure; checking the handedness of the conformer; aligning the conformer to the high resolution structure; and performing a van der Waals calculation.

[0017] Still another embodiment of the present invention is a computer-assisted method for use in modifying a protein comprising the steps of: inputting residues numbers of the flexible loop of a protein into a computer-assisted modeling program; inputting parameters for analysis into a computer-assisted modeling program; general $\phi$, $\psi$, $\omega$ angles randomly in allowed region of Ramachandran maps; generating backbone atoms; performing a CA-CA distance check for the N and C termini; minimizing the N and C termini of the conformer to be the same as the high resolution structure; checking the handedness of the conformer; performing van der Waals check on backbone atoms; aligning the conformer to the high resolution structure; performing van der Waals check on backbone and template protein atoms; adding sidechains to the backbone atoms; and performing van der Waals check on all atoms.

[0018] An embodiment of the present invention is a method for determining the rate of disulfide bond loop closure in a protein comprising at least one two-cysteine motif represented by $C-X_n-C$ where n is an integer, the method comprising the steps of: performing a van der Waals calculation on a multiplicity of conformers of the peptide and subtracting those conformers that cannot form an intramolecular disulfide to yield an ensemble of $N_o$ sterically allowed conformers; analyzing the ensemble of sterically allowed conformers to yield an ensemble of $N_c$ conformers that can potentially form an intramolecular disulfide bond; and calculating the ratio $N_c/N_o$ which represents the rate of disulfide bond loop closure in the peptide. In another embodiment the rate may be compared to the rate of disulfide-bond loop closure of the peptide containing at least one different two-cysteine motif. In a further embodiment, the method may comprise the step of generating peptide backbone coordinates for the $C-X_n-C$ motif from standard bond angles, bond lengths and $\phi$, $\psi$, $\omega$ dihedral angles

randomly obtained within the allowed regions of $\phi$, $\psi$ Ramachandran map for each residue to yield the multiplicity of conformers of the peptide. In a further embodiment, the method may further comprise the step of using a side chain rotamer library to generate $C-X_n-C$ side chain coordinates to yield the multiplicity of conformers of the peptide. In another embodiment of the present invention, analyzing the sterically allowed conformers may comprise calculating the free energy of the conformers based upon the solvent accessible surface area. In yet another embodiment of the present invention, analyzing the sterically allowed conformers may further comprise flexibly modeling the cysteine side chains. In a further embodiment of the present invention, the method may comprise the step of weighting $N_c$ and $N_o$ by the difference in free energy ($\Delta G$) between the dithiol and disulfide forms of the $C-X_n-C$ motif and calculating the ratio

$$\sum_{i=0}^{N_C} e^{-\Delta G_i / RT} \Big/ \sum_{i=0}^{N_O} e^{-\Delta G_i / RT}$$

which represents the energy-weighted rate of disulfide loop closure in the peptide. In a further embodiment of the present invention, the method may comprise the step of identifying an ensemble of $N_c$ conformers of the protein that can potentially form an intramolecular disulfide bond. In another embodiment of the present invention, docking the ensemble of $N_c$ conformers to a binding site on a template biomolecule to yield an ensemble of aligned conformers; and performing a van der Waals calculation on the ensemble of aligned conformers to yield an ensemble of $N_b$ sterically allowed conformers that bind to the template biomolecule. In a further embodiment docking the ensemble of $N_c$ conformers to a binding site on a template biomolecule may comprise the steps of: aligning the ensemble of $N_c$ conformers to a binding site on a template biomolecule to yield an ensemble of aligned conformers; and performing a van der Waals calculation on the ensemble of aligned conformers to yield an ensemble of $N_b$ sterically allowed conformers that bind to the template biomolecule. In another embodiment of the invention, the peptide may further comprise a plurality of two-cysteine motifs represented by $C-X_n-C$ wherein n is independently an integer for each two-cysteine motif.

[0019] Another embodiment of the present invention is a method for assessing the binding affinity of a protein to a template molecule. wherein the protein comprises at least one two-cysteine motif represented by $C-X_n-C$ where n is an integer, the method comprising:

docking the ensemble of $N_c$ conformers to a binding site on a template biomolecule to yield an ensemble of $N_b$ conformers that bind the template biomolecule; and calculating the ratio $N_b/N_c$ which is indicative of the binding affinity of the protein for the template biomolecule.

[0020] Yet another embodiment of the present invention is a method for assessing the binding affinity of a protein to a template molecule, wherein the protein comprises at least one two-cysteine motif represented by $C-X_n-C$ where n is an integer, the method comprising the steps of: screening a population of candidate peptides comprising at least one two-cysteine motif represented by $C-X_n-C$ where n is an integer to yield a plurality of candidate peptides that can potentially form an intramolecular disulfide bond; and performing the method of docking the ensemble of $N_b$ conformers to a binding site on a template biomolecule to yield an ensemble of $N_b$ conformers that bond the template biomolecule and calculating the ratio $N_b/N_c$ which is indicative of the binding affinity of the protein for the template biomolecule on at least one candidate peptide that can potentially form an intramolecular disulfide bond to assess the binding affinity of the candidate peptide. In another embodiment of the invention each candidate peptide may comprise a pre-selected amino acid sequence. In yet another embodiment of the invention the pre-selected amino acid sequence may predispose the peptide to form a desired secondary structure. In another embodiment of the invention the desired secondary structure may be a β-turn.

[0021] Another embodiment of the invention is a method for modifying a protein comprising the steps of: evaluating the X-ray crystal structure or a nuclear magnetic resonance solution structure comprising an oxidized reference peptide bound to a target molecule, the reference peptide comprising at least one intramolecular disulfide bond, to identify at least two amino acids at positions favorable to intramolecular disulfide bond formation; substituting cysteines for the two amino acids in the reference peptide to yield a modified peptide comprising at least four cysteines; identifying an ensemble of $N_c$ conformers of the modified peptide that can potentially form at least two intramolecular disulfide bonds; docking the ensemble of $N_c$ conformers to the binding site on the template biomolecule to yield an ensemble of $N_b$ conformers that bind the template bimolecular; calculating the ratio $N_b/N_c$ which is indicative of the binding affinity of the modified peptide for the template biomolecule; and repeating steps (i.)-(v.) to yield modified peptides having cysteine substitutions at different positions so as to identify modified peptides with the highest $N_b/N_c$ ratios. In another embodiment of the invention identifying an ensemble of $N_c$ conformers of the modified peptide that can potentially form at least two intramolecular

6

disulfide bonds may comprise the steps of: identifying a first conformer of the peptide that can potentially form a first intramolecular disulfide bond defining a first disulfide-bonded loop; constraining the model by the first disulfide bond; and identifying a second conformer of the peptide that can potentially form a second intramolecular disulfide bond defining a second longer disulfide-bonded loop. In another embodiment of the invention, if a second conformer is not identified after about 5 to about 10 attempts to identify said conformer, the method may further comprise the steps of: eliminating the first disulfide bond from the model; identifying a first conformer of the peptide that can potentially form a first intramolecular disulfide bond defining a different first disulfide-bonded loop; constraining the model by the first disulfide bond; and identifying a second conformer of the peptide that can potentially form a second intramolecular disulfide bond defining a second longer disulfide-bonded loop.

[0022] Another embodiment of the present invention is a method for assessing the binding affinity of a protein to a template molecule, wherein the protein comprises a flexible loop, the method compromising the steps of: generating a peptide conformation of length N from a starting residue I and matching to a target residue $I + N$ on the peptide model; accepting the loop conformation when the deviation between residue N and the target residue is small; closing the loop using a geometric minimization method; selecting the residue conformation by the method of performing a van der Waals calculation on a multiplicity of conformers of the peptide and subtracting those conformers that cannot form an intramolecular disulfide to yield an ensemble of $N_o$ sterically allowed conformers; analyzing the ensemble of sterically allowed conformers to yield an ensemble of $N_c$ conformers that can potentially form an intramolecular disulfide bond; and calculating the ratio $N_c/N_o$ which represents the rate of disulfide bond loop closure in the peptide and generating peptide backbone coordinates for the $C-X_n-C$ motif from standard bond angles, bond lengths and $\phi$, $\psi$, $\omega$ dihedral angles randomly obtained within the allowed regions of a $\phi$, $\psi$ Ramachandran map for each residue to yield the multiplicity of conformers of the peptide; generating an ensemble of surface loops; and estimating the binding affinity by testing the docking of the full mini-protein ensemble and peptide target containing the loop ensemble.

[0023] Another embodiment of the invention is a protein produced by a computer-assisted method for use in modifying a protein comprising the steps of: inputting a peptide sequence and parameters for analysis into a computer-assisted modeling program; randomly generating $\phi$, $\psi$, $\omega$; generating a peptide backbone; performing a van der Waals check of the

7

backbone; calculating a solvent accessible surface based energy of all conformers; modeling the disulfide(s); performing a van der Waals check after each rotamer is added; and calculating a solvent accessible surface based energy of all conformers.

[0024] An embodiment of the present invention is a protein produced by protein miniaturization comprising modeling a protein to have the necessary active site conformation using the method of inputting a peptide sequence and parameters for analysis into a computer-assisted modeling program; randomly generating $\phi$, $\psi$, $\omega$; generating a peptide backbone; performing a van der Waals check of the backbone; calculating a solvent accessible surface based energy of all conformers; modeling the disulfide(s); performing a van der Waals check after each rotamer is added; and calculating a solvent accessible surface based energy of all conformers while reducing the total number of amino acids in the protein.

[0025] Another embodiment of the present invention is a protein capable of docking into a binding site wherein the conformation of a portion of said protein was constrained by the introduction of a disulfide bond by the method of inputting a peptide sequence and parameters for analysis into a computer-assisted modeling program; randomly generating $\phi$, $\psi$ and $\omega$; generating a peptide backbone; performing a van der Waals check of the backbone; calculating a solvent accessible surface based energy of all conformers; modeling the disulfide(s); performing a van der Waals check after each rotamer is added; and calculating a solvent accessible surface based energy of all conformers.

[0026] An embodiment of the present invention is a protein, created by the method of inputting a peptide sequence and parameters for analysis into a computer-assisted modeling program; randomly generating $\phi$, $\psi$ and $\omega$; generating a peptide backbone; performing a van der Waals check of the backbone; calculating a solvent accessible surface based energy of all conformers; modeling the disulfide(s); performing a van der Waals check after each rotamer is added; and calculating a solvent accessible surface based energy of all conformers, having the characteristic of inhibiting the binding of a virus to a cell wherein the protein is based upon a tertiary structure of a toxin and comprises at least one loop constrained by a disulfide. In still another embodiment, and ensemble of intramolecular disulfide bond-forming conformers of said loop from the protein may be produced by this method.

[0027] Still another embodiment of the present invention is a protein having decreased conformational entropy loss upon binding to a template molecule in comparison to the naturally occurring protein due to the constraint of at least one loop of an unstable region of a protein in conformational space by the formation of a disulfide and other than disulfide

8

bonds found in the naturally occurring protein using the method of inputting a peptide sequence and parameters for analysis into a computer-assisted modeling program; randomly generating $\phi$, $\psi$ and $\omega$; generating a peptide backbone; performing a van der Waals check of the backbone; calculating a solvent accessible surface based energy of all conformer; modeling the disulfide(s); performing a van der Waals check after each rotamer is added; and calculating a solvent accessible surface based energy of all conformers. In still another embodiment, and ensemble of intramolecular disulfide bond-forming conformers of said loop of the protein may be produced by this method.

[0028] An embodiment of the present invention is a protein produced by a computer-assisted method for use is modifying a protein comprising the steps of: inputting a peptide sequence and parameters for analysis; searching conformational space in the allowed regions of the Ramachandran plots; minimizing the N and C termini of the conformer to be the same as the high resolution structure; checking the handedness of the conformer; aligning the conformer to the high resolution structure; and performing a van der Waals calculation.

[0029] Another embodiment of the present invention is a protein modified by the method comprising the steps of: evaluating a X-ray crystal structure or a nuclear magnetic resonance solution structure comprising an oxidized reference peptide bound to a target molecule, the reference peptide comprising at least one intramolecular disulfide bond, to identify at least two amino acids at positions favorable to intramolecular disulfide bond formation; substituting cysteines for the two amino acids in the reference peptide to yield a modified peptide comprising at least four cysteines; identifying an ensemble of $N_c$ conformers of the modified peptide that can potentially form at least two intramolecular disulfide bonds; docking the ensemble of $N_c$ conformers to the binding site on the template biomolecule to yield an ensemble of $N_b$ conformers that bind the template biomolecule; calculating the reaion $N_b/N_c$ which is indicative of the binding affinity of the modified peptide for the template biomolecule; and repeating these steps to yield modified peptides having cysteine substitutions at different positions so as to identify modified peptides with the highest $N_b/N_c$ ratios.

[0030] A further aspect of the invention is a computer system that can implement the described methods. The computer system has a software program coded to perform the described methods. Preferably, a software program would read protein sequence data from a database or from an input file. One embodiment of such a computer system for designing a modified-protein includes a database containing a set of protein sequence data and a software

9

program coupled with said database for interaction with the database. The software program is adapted for performing the steps of generating randomly conformational angles from the set of protein sequence data, generating a protein backbone using the conformational angles, performing a van der Waals calculation of the protein backbone, calculating a solvent accessible surface based energy of conformers, modeling disulfide bonds in the protein backbone, performing a van der Waals calculation for the disulfide bonds, calculating a solvent accessible surface based energy of conformers that are generated in previous steps, and creating the modified protein with structural characteristics found in the above steps.

[0031] Another embodiment of a computer system for designing a modified-protein has a database containing a set of protein sequence data and a software program coupled with said database. The software program is adapted for performing the steps of generating randomly conformational angles in allowed region of Ramachandran maps from the set of protein sequence data, generating a protein backbone using said conformational angles, determining disulfide bonds in the protein backbone, calculating linear conformers, calculating solvent accessible surface based energy of conformers that are generated in previous steps, and creating the modified protein using structural characteristics identified in the above steps. The calcluating step may be performed by the software or linked to an external program for calculating conformers.

[0032] Another aspect of the invention is a computer-readable storage medium having stored therein a software program that is capable of executing the methods described herein. The computer-readable medium may be any storage-readable medium utilized by a computer, for purposes of illustration but not for limitation, may include floppy disks, hard drives, storage drives, disk packs, ROM, RAM, PC cards, optical media, and magnetic media.

[0033] Other objects, features and advantages of the present invention will become apparent from the following detailed description. It should be understood, however, that the detailed description and the specific examples, while indicating preferred embodiments of the invention, are given by way of illustration only, since various changes and modifications within the spirit and scope of the invention will become apparent to those skilled in the art from this detailed description.

## BRIEF SUMMARY OF THE DRAWINGS
[0034] The following drawings form part of the present specification and are included to further demonstrate certain aspects of the present invention. The invention may be better

understood by reference to one or more of these drawings in combination with the detailed description of specific embodiments presented herein.

[0035] FIG. 1. In the diagram, the letters A, B and C are the starting positions of the three atoms. The letters p, q and v are the vectors whose lengths are the standard bond lengths. X0, a temporary position of X, is in the q direction. The distance between C and X0 is same as the C-X bond length.

[0036] FIG. 2. Two successive (residue i-1 and i) peptide units are selected from the polypeptide backbone. Rotation about the N-Cα bond is denoted by φ, and rotation about Cα-C bond by ψ and about C-N bond by ω.

[0037] FIG. 3A, FIG. 3B, FIG. 3C and FIG. 3D. Assignment of φ and ψ angles. The less-favored regions, bounded by dashed lines, are given 30% point pairs less than the favored regions that are bounded by solid lines. The pictures are respectively FIG. 3A, for ALA, FIG. 3B, for VAL, FIG. 3C for GLY, FIG. 3D for PRO.

[0038] FIG. 4. The diagram demonstrates the two cysteines (i and j) that are assigned for disulfide modeling. The positions of the sulfur SG are restricted on the circles that are obtained by the rotation along the Cα-Cβ bond.

[0039] FIG. 5. The MPMOD flow chart. The input parameters (step 101), such as the peptide sequence and disulfide bond connectivity, are loaded, then the conformational angles (φ, ψ, ω) are generated in the four maps (step 102). The atoms of main chain and side chain are generated base on the angles. The van der Waals checks are performed separately for the backbone atoms and side chain atoms (step 103). If there is a Van der Waals violation, the conformer will be rejected. It will go back to get another set of conformational angles until the peptide is finished without any atom collisions. Then the coordinates of peptide are recorded and the solvent accessible surface (SAS) based energy is calculated (step 104). The disulfide bond is modeled to see if there is a disulfide bond is possible for the two residue pairs (step 105). If a disulfide bond is possible, the SAS energy for this conformer is calculated. If a disulfide bond is not possible, another set of conformational angles is tried and the procedure is repeated until a conformer with a disulfide bond is obtained. Finally, the SAS energy is calculated for this conformer again (step 106).

[0040] FIG. 6A and FIG. 6B. Comparison of probabilities obtained from modeling with the equilibrium constant $K_d$ obtained from experiment (Zhang & Snyder, 1989). The probabilities have been scaled to the experimental values. The scale factor for CXC and CXXC was defined by $K=\Sigma(Exp)/\Sigma(Mod)$, where $\Sigma(Exp)$ is the sum of all the experimental

11

values and $\Sigma(\text{Mod})$ is the sum of all modeled values. Each individual probability of the modeled conformer is multiplied by the scale factor K to get the scaled value. The dark bars are from experiment and the light ones are from calculation. FIG. 6A shows the values for CXC. FIG. 6B shows the values for CXXC.

[0041] FIG. 7. The variation of the ratio $N_c/N_o$ as the number of conformer increases in the ensemble. Here, only the CXXC series is given. The CXC series is also similar.

[0042] FIG. 8. Comparison of probabilities obtained from modeling with the equilibrium constant $K_c$ obtained from experiment (Zhang & Snyder, 1989). Series 1 is the ratio of number of hydrogen bonds in the SS bond closed conformers divided by the total number of conformer. Series 2 is the ratio weighted by the state probability

$$\sum_{i=0}^{N_c} e^{-\Delta G_i / RT} \Big/ \sum_{i=0}^{N_o} e^{-\Delta G_i / RT}$$

Series 3 is from the $K_c$ from experiment. Series 4 is the ratio $N_c/N_o$ from the hard sphere atom model. In order to put them into one picture, we scaled all the series in following factors: series 1 (*1000), series 1 (*10000), series 2 (*100), series 3 (*0.1), series 4 (*1000).

[0043] FIG. 9. Interactions for the peptide-streptavidin complex. Here the peptide has two disulfide bonds that are cross-linked. The HPQ motif is sitting in the binding pocket and there are three hydrogen bonds involving in the interaction for the complex.

[0044] FIG. 10. Disulfide-bonded random conformations for the ensemble CCHPQCGMVEEC. Each conformer has two cross-linked disulfide bonds. The randomly generated conformer has various conformations.

[0045] FIG. 11. The number of chances for each residue of the peptide CCHPQCGMVEEC to collide with the target streptavidin.

[0046] FIG. 12. Correlation of the "binding ratio" with the observed binding constant $K_a$. The straight line is fitted by minimizing the summation Res $= \sum_{i=1}^{N} (\Delta G_f - \Delta G_m)^2$.

[0047] FIG. 13A and FIG. 13B. Flow chart of the MPMOD program (Fast Mod).

[0048] FIG. 14A and FIG. 14B. Flow chart of the MPMOD program (Slow Mod).

[0049] FIG. 15. Flow chart of the MPMOD program (Loop Generation).

[0050] FIG. 16. Flow chart of the MPMOD program's modeling of disulfide bonds.

[0051] FIG. 17. Flow chart of the MPMOD program's binding test.

## DETAILED DESCRIPTION OF THE INVENTION

[0052] The present application includes methods of modifying a peptide to increase the binding affinity of a template molecule by increasing the stability of a peptide by decreasing the conformational entropy loss upon binding to the template molecule.

### I.      Definitions

[0053] A or an, as used herein in the specification, may mean one or more than one. As used herein in the claim(s), when used in conjunction with the word "comprising", the words "a" or "an" may mean one or more than one.

[0054] Another, as used herein, may mean at least a second or more.

[0055] **Based upon a tertiary structure**, as used herein, refers to a structure that possesses a similar backbone structure to that of the original structure that it is referred to being based upon.

[0056] **Conformer**, as used herein, refers to various non-superimposable three-dimensional arrangements of atoms that are interconvertible without breaking covalent bonds.

[0057] **Constrained**, as used herein, refers to a limitation in the conformational space that the peptide may adopt.

[0058] **Disulfide bridge** and **disulfide bond** as used herein, refers to a covalent bond between the sulfur atoms of two cysteines.

[0059] **Generate**, as used herein, refers the act of defining or originating by the use of one or more operations.  The individuals using the invention may create the matter or data themselves or locate the matter or data elsewhere and utilize it in the practice of the invention.

[0060] **Loop**, as used herein, are turns in the polypeptide chain that reverse the direction of the polypeptide chain at the surface of the molecule.

[0061] **Rotamer**, as used herein, refers to a low energy amino acid side chain conformation.

[0062] **Peptide**, as used herein, refers to a chain of amino acids with a defined sequence whose physical properties are those expected from the sum of its amino acid residues and there is no fixed three-dimensional structure.

[0063] **Protein**, as used herein, refers to a chain of amino acid residues usually of defined sequence, length and three-dimensional structure. The polymerization reaction which produces a protein results in the loss of one molecule of water from each amino acid, proteins

13

are often said to be composed of amino acid residues. Natural protein molecules may contain as many as 20 different types of amino acid residues, each of which contains a distinctive side chain. A protein may be composed of multiple peptides.

[0064] **Structural Characteristics**, as used herein, refers to the characteristics that are determined using the computer-assisted program, such as, but not limited to folding characteristics, disulfide bonding, binding affinity, aggregation, solubility, immunogenicity, stablility, etc. Thus, one of skill in the art realizes that the present invention is used to determine any structural characteristic of a protein and this characteristic may be enhanced or reduced depending upon the application of use.

[0065] **Template molecule**, as used herein, refers to the protein to which the modified protein is binding.

## II.    MPMOD

[0066] Combination of the random search of the conformational space in the allowed regions of Ramachandran plot, using the simple hard sphere model to generate the stereo-chemically acceptable conformers, and a flexible disulfide bond modeling, provides a simple and useful tool to study the behavior of cyclic peptides. The "rate" or probability of SS bond loop closure as defined by $N_c/N_o$ converges to a stable value when the ensemble has more than 1000 conformer. For the CXC and CXXC series of peptide, the modeled probability of loop closure behaves the same way as the experimentally determined equilibrium constant $K_c$ for all the four types of the peptides. Both compare well after a common scale factor is applied. The geometry or van der Waals interaction plays a dominant role for the loop closure for the small peptides CXC and CXXC. One of skill in the art realizes that the MPMOD method of protein design is not limited to protein pharmaceuticals. For example, it includes, but is not limited to, the use of the MPMOD method to design proteins that may be beneficial as a diagnostic reagents, research reagents, pesticides or herbicides.

[0067] The program (MPMOD) is an efficient method to generate disulfide-bonded conformers. It takes about 10~20 CPU minutes to obtain 4000 disulfide bonded conformers CXXC using a Linux system on a Pentium III 450 MHz. Because the conformer CXC has higher probability of collision, it takes about 3 times more CPU time than to generate the CXXC. However, the consumed CPU time strongly depended on the criteria used to generate the conformer.

14

[0001] The basic MPMOD program comprises the following steps. The input parameters, such as the peptide sequence and disulfide bond connectivity, are loaded, then the conformational angles ($\phi$, $\psi$, $\omega$) are generated in the four maps. The atoms of main chain and side chain are generated base on the angles. The van der Waals checks are performed separately for the backbone atoms and side chain atoms. If there is a Van der Waals violation, the conformer will be rejected. It will go back to get another set of conformational angles until the peptide is finished without any atom collisions. Then the coordinates of peptide are recorded and the solvent accessible surface (SAS) based energy is calculated. The disulfide bond is modeled to see if there is a disulfide bond is possible for the two residue pairs. If a disulfide bond is possible, the SAS energy for this conformer is calculated. If a disulfide bond is not possible, another set of conformational angles is tried and the procedure is repeated until a conformer with a disulfide bond is obtained. Finally, the SAS energy is calculated for this conformer again.

[0069] MPMOD can be used to generate disulfide bonded conformers and/or linear conformers. To generate linear conformers in conjuction with disulfide bonded conformers MPMOD is linked to the COREX algorithm (Hisler & Freire, 1996).

## III.    Prokaryotic Peptide Display

[0070] Molecular analysis of naturally occurring and artificial protein libraries has been greatly improved by the development of various "display" methodologies. The general scheme behind display techniques is the advantageous expression of peptides, and their disposition on some biological surface (phage, cell, etc.). The ability of different versions of the displaying organism to present millions and millions of different variants allows the rapid screening of the corresponding library for biological function.

[0071] In U.S. Patent 5,821,047, monovalent phage display is described. This method provides for the selection of novel proteins, and variants thereof. The method comprises fusing a gene encoding a protein of interest to the carboxy terminal domain of the gene III coat protein of the filamentous phage M13. The fusion is mutated to form a library of structurally related fusion proteins that are expressed in low quantity on the surface of phagemid candidates.

[0072] U.S. Patent 5,571,698 describes directed evolution using an M13 phagemid system. A protein is expression as a fusion with the M13 gene III protein. Successive rounds

15

of mutagenesis are performed, each time selecting for improved biological function, *e.g.,* binding of a protein to a cognate binding partner.

[0073] Heterodimer phage libraries are described in U.S. Patent 5,759,817. Filamentous phage comprising a matrix of cpVIII proteins encapsulating a genome encoding first and second polypeptides of an autogenously assembling receptor, such as an antibody, are provided. The receptor is surface-integrated into the phage coat matrix via the cpVIII membrane anchor, presenting the receptor for biological assessment.

[0074] Another system, lambdoid phage, also can be used for display purposes. In U.S. Patent 5,672,024, lambdoid phage comprising a matrix of proteins encapsulating a genome encoding first and second polypeptides of an autogenously assembling receptor are prepared. The surface-integrated receptor is available on the surface on the phage for characterization.

[0075] Immunoglobulin heavy chain libraries are displayed by phage as described in U.S. Patent 5,824,520. A single chain antibody library is generated by creating highly divergent, synthetic hypervariable regions, followed by phage display and selection. The resulting antibodies were used to inhibit intracellular enzyme activity. Another patent describing antibody display is U.S. Patent 5,922,545.

[0076] Another example of phage display can be found in U.S. Patent 5,780,279. This method provides for the identification and selection of novel substrates for enzymes. The method comprises constructing a gene fusion comprising DNA encoding a polypeptide fused to a DNA encoding a substrate peptide, which in turn is fused to DNA encoding at least a portion of a phage coat protein. The DNA encoding the substrate peptide is mutated at one or more codons, thereby generating a family of mutants. The fusion protein is expressed on the surface of the phagemid particle and subjected to chemical or enzymatic modification of the substrate peptide. Those phagemid particles that have been modified are then separated from those that have not.

[0077] Bacteria also have been used successfully to display proteins. U.S. Patent 5,348,867, describes expression of proteins on bacterial surfaces. The compositions and methods provide stable, surface-expressed polypeptide from recombinant gram-negative bacterial cell hosts. A tripartite chimeric gene and its related recombinant vector include separate DNA sequences for directing or targeting and translocating a desired gene product from a cell periplasm to the external cell surface. A wide range of polypeptides may be

16

efficiently surface expressed using this system. See also, U.S. Patents 5,508,192 and 5,866,344.

[0078] U.S. Patent 5,500,353 describes another bacterial display system. Bacteria (*e.g., Caulobacter*) having a S-layer modified such that the bacterium S-layer protein gene contains one or more in-frame fusions coding for one or more heterologous peptides or polypeptides is described. The proteins are expressed on the surface of the bacterium, which may advantageously be cultured as a film.

## IV.    Mutagenesis

[0079] Where employed, mutagenesis will be accomplished by a variety of standard, mutagenic procedures. Mutation is the process whereby changes occur in the quantity or structure of an organism. Mutation can involve modification of the nucleotide sequence of a single gene, blocks of genes or whole chromosome. Changes in single genes may be the consequence of point mutations that involve the removal, addition or substitution of a single nucleotide base within a DNA sequence, or they may be the consequence of changes involving the insertion or deletion of large numbers of nucleotides.

[0080] Mutations can arise spontaneously as a result of events such as errors in the fidelity of DNA replication or the movement of transposable genetic elements (transposons) within the genome. They also are induced following exposure to chemical or physical mutagens. Such mutation-inducing agents include ionizing radiations, ultraviolet light and a diverse array of chemical such as alkylating agents and polycyclic aromatic hydrocarbons all of which are capable of interacting either directly or indirectly (generally following some metabolic biotransformations) with nucleic acids. The DNA lesions induced by such environmental agents may lead to modifications of base sequence when the affected DNA is replicated or repaired and thus to a mutation. Mutation also can be site-directed through the use of particular targeting methods.

[0081] Structure-guided site-specific mutagenesis represents a powerful tool for the dissection and engineering of protein-ligand interactions (Wells *et al.*, 1996). The technique provides for the preparation and testing of sequence variants by introducing one or more nucleotide sequence changes into a selected DNA.

[0082] Site-specific mutagenesis uses specific oligonucleotide sequences that encode the DNA sequence of the desired mutation, as well as a sufficient number of adjacent, unmodified nucleotides. In this way, a primer sequence is provided with sufficient size and

17

complexity to form a stable duplex on both sides of the deletion junction being traversed. A primer of about 17 to 25 nucleotides in length is preferred, with about 5 to 10 residues on both sides of the junction of the sequence being altered.

[0083] The technique typically employs a bacteriophage vector that exists in both a single-stranded and double-stranded form. Vectors useful in site-directed mutagenesis include vectors such as the M13 phage. These phage vectors are commercially available and their use is generally well known to those skilled in the art. Double-stranded plasmids are also routinely employed in site-directed mutagenesis, which eliminates the step of transferring the gene of interest from a phage to a plasmid.

[0084] In general, one first obtains a single-stranded vector, or melts two strands of a double-stranded vector, which includes within its sequence a DNA sequence encoding the desired protein or genetic element. An oligonucleotide primer bearing the desired mutated sequence, synthetically prepared, is then annealed with the single-stranded DNA preparation, taking into account the degree of mismatch when selecting hybridization conditions. The hybridized product is subjected to DNA polymerizing enzymes such as *E. coli* polymerase I (Klenow fragment) in order to complete the synthesis of the mutation-bearing strand. Thus, a heteroduplex is formed, wherein one strand encodes the original non-mutated sequence, and the second strand bears the desired mutation. This heteroduplex vector is then used to transform appropriate host cells, such as *E. coli* cells, and clones are selected that include recombinant vectors bearing the mutated sequence arrangement.

[0085] Other methods of site-directed mutagenesis are disclosed in U.S. Patents 5,220,007; 5,284,760; 5,354,670; 5,366,878; 5,389,514; 5,635,377; and 5,789,166.

18

## V.      Modified Polynucleotides and Polypeptides

[0086] Amino acid substitutions are generally based on the relative similarity of the amino acid side-chain substituents, for example, their hydrophobicity, hydrophilicity, charge, size, and/or the like.  An analysis of the size, shape and/or type of the amino acid side-chain substituents reveals that arginine, lysine and/or histidine are all positively charged residues; that alanine, glycine and/or serine are all a similar size; and/or that phenylalanine, tryptophan and/or tyrosine all have a generally similar shape.   Therefore, based upon these considerations, arginine, lysine and/or histidine; alanine, glycine and/or serine; and/or phenylalanine, tryptophan and/or tyrosine; are defined herein as biologically functional equivalents.

[0087] To effect more quantitative changes, the hydropathic index of amino acids may be considered.  Each amino acid has been assigned a hydropathic index on the basis of their hydrophobicity and/or charge characteristics, these are: isoleucine (+4.5); valine (+4.2); leucine (+3.8); phenylalanine (+2.8); cysteine/cystine (+2.5); methionine (+1.9); alanine (+1.8); glycine (-0.4); threonine (-0.7); serine (-0.8); tryptophan (-0.9); tyrosine (-1.3); proline (-1.6); histidine (-3.2); glutamate (-3.5); glutamine (-3.5); aspartate (-3.5); asparagine (-3.5); lysine (-3.9); and/or arginine (-4.5).

[0088] The importance of the hydropathic amino acid index in conferring interactive biological function on a protein is generally understood in the art (Kyte & Doolittle, 1982, incorporated herein by reference).  It is known that certain amino acids may be substituted for other amino acids having a similar hydropathic index and/or score and/or still retain a similar biological activity.  In making changes based upon the hydropathic index, the substitution of amino acids whose hydropathic indices are within ±2 is preferred, those which are within ±1 are particularly preferred, and/or those within ±0.5 are even more particularly preferred.

[0089] It also is understood in the art that the substitution of like amino acids can be made effectively on the basis of hydrophilicity, particularly where the biological functional equivalent protein and/or peptide thereby created is intended for use in immunological embodiments, as in certain embodiments of the present invention.  U.S. Patent 4,554,101, incorporated herein by reference, states that the greatest local average hydrophilicity of a protein, as governed by the hydrophilicity of its adjacent amino acids, correlates with its immunogenicity and/or antigenicity, i.e., with a biological property of the protein.

[0090] As detailed in U.S. Patent 4,554,101, the following hydrophilicity values have been assigned to amino acid residues: arginine (+3.0); lysine (+3.0); aspartate (+3.0 ± 1);

glutamate (+3.0 ± 1); serine (+0.3); asparagine (+0.2); glutamine (+0.2); glycine (0); threonine (-0.4); proline (-0.5 ± 1); alanine (-0.5); histidine (-0.5); cysteine (-1.0); methionine (-1.3); valine (-1.5); leucine (-1.8); isoleucine (-1.8); tyrosine (-2.3); phenylalanine (-2.5); tryptophan (-3.4). In making changes based upon similar hydrophilicity values, the substitution of amino acids whose hydrophilicity values are within ±2 are preferred, those that are within ±1 are particularly preferred, and/or those within ±0.5 are even more particularly preferred.

## VI. Mimetics

[0091] The present inventors contemplate that structurally similar compounds may be formulated to mimic the key portions of peptide or polypeptides. Such compounds may be termed peptidomimetics.

[0092] Certain mimetics that mimic elements of protein secondary and tertiary structure are described in Johnson et al. (1993). The underlying rationale behind the use of peptide mimetics is that the peptide backbone of proteins exists chiefly to orient amino acid side chains in such a way as to facilitate molecular interactions, such as those of antibody and/or antigen. A peptide mimetic is thus designed to permit molecular interactions similar to the natural molecule.

[0093] Some successful applications of the peptide mimetic concept have focused on mimetics of β-turns within proteins, which are known to be highly antigenic. As discussed herein, possible β-turn structure within a polypeptide can be predicted by computer-based algorithms. Once the component amino acids of the turn are determined, mimetics can be constructed to achieve a similar spatial orientation of the essential elements of the amino acid side chains.

[0094] Other approaches have focused on the use of small, multidisulfide-containing proteins as attractive structural templates for producing biologically active conformations that mimic the binding sites of large proteins. Vita et al. (1998). A structural motif that appears to be evolutionarily conserved in certain toxins is a small (30-40 amino acids), stable, and highly permissive for mutation motif. This motif is composed of a beta sheet and an alpha helix bridged in the interior core by three disulfides.

[0095] Beta II turns have been mimicked successfully using cyclic L-pentapeptides and those with D-amino acids. Weisshoff et al. (1999). Also, Johannesson et al. (1999) report on bicyclic tripeptides with reverse turn inducing properties.

20

[0096] Methods for generating specific structures have been disclosed in the art. For example, alpha-helix mimetics are disclosed in U.S. Patents 5,446,128; 5,710,245; 5,840,833; and 5,859,184. Theses structures render the peptide or protein more thermally stable, also increase resistance to proteolytic degradation. Six, seven, eleven, twelve, thirteen and fourteen membered ring structures are disclosed.

[0097] Methods for generating conformationally restricted beta turns and beta bulges are described, for example, in U.S. Patents 5,440,013; 5,618,914; and 5,670,155. Beta-turns permit changed side substituents without having changes in corresponding backbone conformation, and have appropriate termini for incorporation into peptides by standard synthesis procedures. Other types of mimetic turns include reverse and gamma turns. Reverse turn mimetics are disclosed in U.S. Patents 5,475,085 and 5,929,237, and gamma turn mimetics are described in U.S. Patents 5,672,681 and 5,674,976.

## VII.    EXAMPLES

[0098] The following examples are included to demonstrate preferred embodiments of the invention. It should be appreciated by those skilled in the art that the techniques disclosed in the examples which follow represent techniques discovered by the inventor to function well in the practice of the invention, and thus can be considered to constitute preferred modes for its practice. However, those of skill in the art should, in light of the present disclosure, appreciate that many changes can be made in the specific embodiments which are disclosed and still obtain a like or similar result without departing from the concept, spirit and scope of the invention. More specifically, it will be apparent that certain agents that are both chemically and physiologically related may be substituted for the agents described herein while the same or similar results would be achieved. All such similar substitutes and modifications apparent to those skilled in the art are deemed to be within the spirit, scope and concept of the invention as defined by the appended claims.

### Example 1
### Construction of a polypeptide

[0099] A unit peptide was generated using a rotational matrix $[M]^\theta{}_{\lambda,\mu,\nu}$ (Jeffreys & Jeffreys, 1950) for the effect of a rotation on the coordinates of a point by an angle $\theta$ about an axis through the origin having the direction cosines $\lambda$, $\mu$, $\nu$.

21

$$[M]_{\lambda,\mu,\nu}^{\theta} = \begin{bmatrix} \cos\theta + \lambda^2(1-\cos\theta) & \lambda\mu(1-\cos\theta) - \nu\sin\theta & \lambda\nu(1-\cos\theta) + \mu\sin\theta \\ \lambda\mu(1-\cos\theta) + \nu\sin\theta & \cos\theta + \mu^2(1-\cos\theta) & \mu\nu(1-\cos\theta) - \lambda\sin\theta \\ \lambda\nu(1-\cos\theta) - \mu\sin\theta & \mu\nu(1-\cos\theta) + \lambda\sin\theta & \cos\theta + \nu^2(1-\cos\theta) \end{bmatrix}$$

**Eq.1**

[0100] Giving the coordinates of three successive atoms A, B and C, the three components $\lambda$, $\mu$, $\nu$ in the matrix were determined. The fourth atom X can be generated based on this matrix and the dihedral angle $\chi$(A-B-C-X), the bond angle $\alpha$(B-C-X) and the bond length d(C-X). FIG.1 gives a diagram for building the fourth atom X.

[0101] For the coordinate used in FIG. 1, the position vectors of the three atoms A, B, C are r1, r2, r3. The vectors **p**, **q** and **v** are respectively **p**=r2-r1, **q**=r3-r2, **v**=d*(q/|q|), where d is the C-X bond length. The unit vector **n**=(p×q)/|p×q| is normal to the plane A-B-C formed by atom A, B and C. X0 is the temporary position in the q direction. The position of atom is first rotated to the X position with the rotational axis **n** and rotational angle $\pi-\alpha$, where $\alpha$ is the bond angle of atom B, C and X. The final position of X is obtained by a rotation with the axis q and the dihedral angle $\chi$ (one of the three dihedral angles $\phi$, $\psi$ and $\omega$). Both the rotation for $\pi-\alpha$ and $\chi$ are clockwise looking down the relevant vectors. The angles are positive if they are clockwise rotation and negative if anti-clockwise rotation. The final position of the atom X is expressed as

$$r_X = r_C + [M]_u^{\chi}[M]_n^{\pi-\alpha} \, v. \qquad\qquad \textbf{Eq. 2}$$

where the notation $[M]_a^{\theta}$ has been used for the matrix $[M]_{\lambda,\mu,\nu}^{\theta}$ of Eq. (1) with a as a unit vector in the direction of $\lambda$, $\mu$, $\nu$.

[0100] Starting from a unit peptide, all of other atoms of a polypeptide can be determined by Eq. (2). The backbone atoms (N, C$\alpha$, N, O, HN, HC$\alpha$) were generated making use of the conformational angles ($\phi$,$\psi$, $\omega$) and the standard bond angles and bond lengths (Momany *et al.*, 1975) are listed in Table 1. The backbone atoms on residue i are built by the following parameters in braces; atom $N_i$ by {$N_{i-1}$, $C\alpha_{i-1}$, $C_{i-1}$, $\phi_{i-1}$, $\alpha(C\alpha_{i-1}$-$C_{i-1}$-$N_i$), d($C_{i-1}$-$N_i$)}; $C\alpha_i$ by {$C\alpha_{i-1}$, $C_{i-1}$, $N_i$, $\omega_{i-1}$, $\alpha(C_{i-1}$-$N_i$-$C\alpha_i$), d($N_i$-$C\alpha_i$)}; $C_i$ by {$C_{i-1}$, $N_i$, $C\alpha_i$, $\psi_i$, $\alpha(N_i$-$C\alpha_i$-$C_i$), d($C\alpha_i$- $C_i$)}; $O_i$ by {$C_{i-1}$, $N_i$, $C\alpha_i$, $\psi_i$, $\alpha(C\alpha_i$-$C_i$-$O_i$), d($C_i$-$O_i$)} and $HN_i$ by {$C\alpha_{i-1}$, $C_{i-1}$, $N_i$, $\alpha(C\alpha_{i-1}$-$C_{i-1}$-$N_i$), d($N_i$-$H_i$)}, where $\alpha$ is the bond angle formed by the three atoms in the parenthesis and d is the bond length of two atoms. C$\beta$ and HC$\alpha$ are treated specially because of the tetrahedral geometry with N and C atoms. Both atoms do not depend

22

on the dihedral angles ($\phi$, $\psi$, $\omega$). Of the two possible positions of the C$\beta$ atom, the one corresponding to the L-amino acid residues has been used throughout the studies. Hence, C$\beta_i$ atom of residue i is built by {$N_i$, C$\alpha_i$, $C_i$, 109.5°, d(C$\alpha_i$ - C$\beta_i$)} and HC$\alpha_i$ by {$N_i$, C$\alpha_i$, $C_i$, -109.5°, d(C$\alpha_i$ - HC$\alpha_i$)}. FIG. 2 shows the diagram for two successive unit peptides.

[0103] The side chain atoms of a polypeptide were built in the same way as shown for building the backbone atoms. The bond lengths and bond angles were taken from the published values (Momany *et al.* 1975). There are maximally four types of dihedral angles ($\chi1$, $\chi2$, $\chi3$, $\chi4$) for the 20 amino acids. Surveys of crystallographic structures of proteins and small peptides show that the $\chi$ angles are highly favored (Janin *et al.*, 1978, Benedetti *et al.*, 1983). Ponder and Richards (1987) studied the population of the $\chi$ angles using 2273 residues obtained from 19 protein crystal structures with a resolution higher than 1.8Å and R-factor below 0.18. Ponder and Richards found that the population of the $\chi$ angles was much higher in some preferred values and that the standard deviation was smaller than previously published values. This indicates that a rotamer library can approximately represent the behavior of a side chain. The rotamer library by Ponder and Richards (1987) was used for all the dihedral angles. In order to reduce bias to as little as possible for addition of the rotamers, the rotamers were selected that did not collide with the backbone atoms for each residue. One set of rotamers was picked randomly from all the residues. However, this set of rotamers may have a van der Waals contact violation. Three random tries were given to increase the chance of getting a suitable set rotamers. If none of the three tries satisfies the van der Waals check, the backbone is rejected and the procedure is repeated.

## Example 2
### Determination of the dihedral angles ($\phi$, $\psi$, $\omega$)

[0100] The subroutine ran2.f from "Numerical Recipes"(Press *et al.*, 1986) was used to generate random numbers that uniformly distribute in the range from 0.0 to 1.0. These random values have no sequential correlation and the period is practically infinite. The conformational angles ($\phi$, $\psi$) are assigned to be ax + b where x is the random value $0 \leq x \leq 1.0$, a and b are two constants which are adjusted so that the two angles are restricted to the allowed regions of the Ramachandran plots.

23

**TABLE 1**

| Bond angle (°) | | Bond length (Å) | |
|---|---|---|---|
| $C\alpha_{i-1}$-$C_{i-1}$-$N_i$ | 115.0 | $C_{I-1}$-$N_i$ | 1.325* |
| $C_{i-1}$-$N_i$-$C\alpha_i$ | 121.0 | $N_I$-$C\alpha_i$ | 1.453 |
| $C_i$-$C\alpha_i$-$N_i$ | 109±5 | $C_I$-$C\alpha_i$ | 1.530 |
| $C\alpha_i$-$C_I$-$O_i$ | 120.5 | $C_I$-$O_i$ | 1.230 |
| $N_i$-$C\alpha_i$-$C\beta_i$ | 109.5 | $C\alpha_i$-$C\beta_i$ | 1.530 |
| $N_i$-$C\alpha_i$-$H_i$ | 109.5 | $C\alpha_i$-$H_i$ | 1.020 |
| $C_{i-1}$-$N_i$-$H_i$ | 121.0 | $N_I$-$H_i$ | 0.940 |

Table 1. The bond lengths and bond angles used for building polypeptide. The bond angle C-Cα-N is not as rigid as other bond angles. Therefore it is allowed to vary ±5° around the average value 109°.

* If the residue $i$ is Pro the bond length of $C_{i-1}$-$N_I$ is 1.355Å.

[0105] In order to sample the conformational space efficiently, the backbone dihedral angles ($\phi$, $\psi$) of a protein are divided into four categories, one for glycine, one for proline, one for the CB-branched amino acids (VAL, ILE and THR), and one for all other amino acids. Glycine, with one hydrogen atom as its side chain, can adopt a wide range of conformations and the map is symmetrical due to absence of an R substitute on the alpha carbon (Cα). Proline only adopts a very narrow range of conformation space because of the pyrrolidine ring attached to the N and Cα atoms, which restrains the conformation greatly. Alanine is a prototype L-amino acid whose conformational space can approximately represent that of other amino acids except for proline and glycine. However, due to the two branches on the CB atoms, the amino acids (VAL, ILE and THR) have more restricted conformational space than ALA (Scheraga, 1992; Chakrabarti & Pal, 1998). FIG. 3 shows distribution of the conformational angles on the four maps. As the number of the random values becomes sufficiently large, the points will evenly distribute in the allowed regions.

[0106] MacArthur & Thornton (1993) made an analysis of the conformational angles for proline residues from non-homologous X-ray crystal structures determined to 2.5Å or better. They showed that majority of the conformation angles are clustered about the mean values of $\phi$, $\psi$= -61°, -35° for the α region and $\phi$, $\psi$= -65°, 150° for the β region. The early

computations (e.g.,Ramachandran et al., 1963; Nemethy and Scheraga, 1965) with hard-sphere potential, a good first order approximation, showed that the region around the $\psi=0°$ was not allowed. When realistic potential was used, this region became partially allowed. Based on this consideration, the inventors divided the map into the favored regions which are bounded by solid lines and less-favored regions which are bounded by dashed lines. It was determined to give the less-favored regions 60-80% of a chance of occurrence less than the favored regions. For all amino acids except glycine and proline, only 5% of random values were assigned to the $\phi$ positive regions. This conformation assignment is similar to the one proposed by Sowdhamini et al. (1993) with differences of the following two aspects: 1) the inventors sampled conformational space to be closer to the Ramachandran plots and two small areas were added to the positive $\phi$ region and 2) the angle distribution for each map is "non-even". These aspects speed up the modeling of the peptide significantly.

[0107] The trans and cis forms of the peptides have dihedral angles of $\omega=180°$ and $0°$. The non-Pro amino acids are favored in trans form by a ratio of approximately 1000:1. With proline, the trans form is only favored by 4:1. Therefore the non-Pro amino acids were set as 100% trans form and the proline was given up to 5% (or optional) cis form and 95% trans form. The dihedral angle $\omega$ is also allowed to have a fluctuation about $5°$ around the value 180 or $0°$.

### Example 3
### van der Waals steric contacts

[0108] The hard sphere atom model was assumed as the scoring function to eliminate grossly improbable conformers. Ramachandran (1963) used X-ray data to determine a list of contact distances for each kind of atom pair (see the "normal" and "extreme" distances of Table 2.) occurring in proteins. These distances are about 0.3 ~ 0.5 Å smaller than the summation of the van der Waals radii of two atoms (Gavezzotti, 1983). Gavezzoti concluded that the structure was less stable for the distance of "extreme limit " than for the distance of "normal limit". However, the short contact distances are usually in the extreme limit if there are hydrogen bonds or other attractive effects. Iijima et al (1987) calibrated the van der Waals radii of atoms using an inverse Ramachandran plot. The calibrations were based on the comparison of the Ramachandran plots obtained from high resolution X-ray data of proteins and peptides with the allowed conformational space for the di-peptide molecular models built from the published standard bond angles and bond lengths. The calibrated contact distances for each atom pair are about 0.1 to 0.2Å shorter than the "extreme limit" (Table 2).

25

[0109] Out of the three kinds of the short contact distances, both the "extreme limit" and the "calibrated" distance were used for the van der Waals check, unless otherwise specified. The "extreme" distances were used only for the backbone atom pairs checked. The "calibrated" distances were used for the side chain to side chain or side chain to backbone atom pair check. The reasons to use two contact distances were the following: (1) to give some flexibility for the backbone and slightly more flexibility for the side chains, (2) to compensate for the inaccuracy that is caused by the fixed geometrical parameters used to build the polypeptides, (3) to include hydrogen bonds or some attractive features in the conformer. (4) to save computing time especially when the side chain atom pairs are involved in van der Waals checks. It should be noted that for each atom, all the possible non-bonded atom pairs are checked. Atom pairs in the same residue are not checked.

**TABLE 2**

|        | Normal (Å) | Extreme (Å) | Calibrated (Å) |
|--------|------------|-------------|----------------|
| H...H  | 2.0        | 1.9         | 1.9            |
| H...O  | 2.4        | 2.2         | 2.2            |
| H...N  | 2.4        | 2.2         | 2.3            |
| H...C  | 2.4        | 2.2         | 2.3            |
| O...O  | 2.7        | 2.6         | 2.4            |
| O...N  | 2.7        | 2.6         | 2.47           |
| O...C  | 2.8        | 2.7         | 2.55           |
| N...N  | 2.7        | 2.6         | 2.54           |
| N...C  | 2.9        | 2.8         | 2.62           |
| C...C  | 3.0        | 2.9         | 2.7            |
| H...S  | 2.5        | 2.3         | 2.3            |
| O...S  | 2.9        | 2.8         | 2.7            |
| N...S  | 2.95       | 2.85        | 2.77           |
| C...S  | 3.05       | 2.95        | 2.85           |
| S...S  | 3.1        | 3.0         | 3.0            |

Table 2 The short contact distances between each kind of atom pair. The normal and extreme distance is from Ramachandran *et al.*, (1963). The calibrated distance is from Iijima *et al.*, (1987).

26

## Example 4
## Modeling disulfide bonds

[0100] The present invention includes a flexible method to search for the potentially existing disulfide bonds in a structure. For the specified cysteine i, the coordinates of the sulfur atom $S^i$ are generated using the torsion angle $\chi^1$(N-C$\alpha$-C$\beta$-S) and the coordinates of atoms $N_i$, $C\alpha_i$ and $C\beta_i$. The positions of $S^i$ atoms must be on a circle which is formed by rotating about the $C\alpha^i$-$C\beta^i$ bond with the rotational angle $\chi^1$. Statistics shows that the favored dihedral angles $\chi^1$ are around –60, 60 and 180°. In the present invention, a wider region around each angle was scanned *i.e.*, $\chi^1$ is from -20 to –100, 20 to 100 and 140 to 220. The coordinates of $S^i$ are recorded every four degrees when rotating about the $C\alpha^i$-$C\beta^i$ bond. The same procedure is applied to the specified residue j. The distances are checked for all the generated atom pairs $S^i$ and $S^j$ on both circles. FIG.4 shows one of the generated sulfur pairs. If the distance between $S^i$ and $S^j$ is within 2.04±0.4Å and the bond angle $C\beta^i$-$S^i$-$S^j$ and $C\beta^j$-$S^j$-$S^i$ within 104±5° and the dihedral angle $\chi^{S\text{-}S}$($C\beta^i$-$S^i$-$S^j$-$C\beta^j$) within |90|±40°, it was assumed that these two cysteines could form a disulfide bond. To ensure that the generated S atom positions are in good geometry with all other atoms of the conformer, a van der Waals check was also performed. This rejects many position pairs, especially when $C\alpha^i$-$C\beta^i$ bond of the first Cys is approximately in line with $C\alpha^j$-$C\beta^j$ of the second Cys residue. The best position pair was then selected.

[0111] To test the procedure for predicting the disulfide bond, 19 disulfide bonds were examined that were not successfully modeled by Sowdhamini *et al.*, (1989). Table 3 lists the data from the modeled and crystallographically observed disulfide bonds. All the disulfide bonds were successfully predicted (Table 3). However, two disulfide bonds have to be modeled by adjusting the criterion of the torsion angle $\chi^{S\text{-}S}$ to be beyond |90|±40°.

27

## TABLE 3

| Protein | d(S-S) | $\chi^{S-S}$ | $\chi^{i-1}$ | $\chi^{j-1}$ | $\chi^{i-2}$ | $\chi^{j-2}$ | $C\beta^i-S^i-S^j$ | $C\beta^j-S^j-S^i$ | |
|---|---|---|---|---|---|---|---|---|---|
| Azurin | 1.98 | -82 | -65 | -57 | -69 | -72 | 103 | 102 | |
| (2AZA) 3B-26B | 1.95 | -87 | -66 | -60 | -66 | -70 | 101 | 102 | |
| Carboxypeptidase | 1.89 | 93 | -79 | -74 | 144 | -57 | 103 | 105 | |
| (5CPA) 138-161 | 1.93 | 92 | -84 | -78 | 138 | -47 | 109 | 99 | |
| Lysozyme (HEW) | 2.11 | 80 | -35 | -178 | 148 | 42 | 105 | 104 | |
| (6LYZ) 76-94 | 2.32 | 70 | -28 | -180 | 153 | 44 | 101 | 107 | |
| Lysozyme (Human) | 2.08 | 95 | 62.6 | -71 | 81 | -58 | 103 | 107 | |
| (1LZ1) 65-81 | 2.30 | 94 | 62 | -66 | 82 | -58 | 104 | 102 | |
| Ovomycoid. third domain | 1.99 | 99 | -179 | -70 | 65 | -47 | 105 | 104 | |
| (2OVO) 24-56 | 1.99 | 105 | -177 | -66 | 64 | -53 | 104 | 103 | |
| Rat mast cell protease | 2.02 | -88 | -81 | -71 | -153 | -86 | 103 | 108 | |
| (3RP2) 42A-58A | 2.20 | -86 | -84 | -66 | -155 | -85 | 100 | 106 | |
| 168A-182A | 1.99 | 159 | -52 | -92 | 82 | -50 | 103 | 113 | |
| | 2.02 | 168 | -48 | -81 | 94 | -73 | 101 | 106 | a |
| Glutathione reductase | 2.06 | -133 | 178 | -32 | 80 | 118 | 101 | 92 | |
| (3GRS) 58-63 | 2.04 | -136 | 178 | -31 | 78 | 120 | 100 | 92 | a |
| Phospholipase A-2 | 2.08 | 108 | -64 | 176 | -54 | 72 | 101 | 100 | |
| (1BP2) 61–91 | 2.01 | 110 | -64 | 178 | -56 | 71 | 101 | 100 | |
| b-Trypsin | 1.98 | -77 | -98 | -78 | -139 | -82 | 105 | 109 | |
| (1TPP) 42-58 | 2.16 | -72 | -108 | -70 | -144 | -81 | 103 | 108 | |

## TABLE 3

| Protein | d(S-S) | $\chi^{S\text{-}S}$ | $\chi^{i-1}$ | $\chi^{j-1}$ | $\chi^{i-2}$ | $\chi^{j-2}$ | $C\beta^i\text{-}S^i\text{-}S^j$ | $C\beta^j\text{-}S^j\text{-}S^i$ |
|---|---|---|---|---|---|---|---|---|
| b-Trypsin complex | 2.03 | -86 | -82 | -73 | -142 | -80 | 107 | 106 |
| (2PTC) 42E-58E | 2.06 | -85 | -82 | -73 | -142 | -79 | 105 | 106 |
| Immunoglobulin FAB | 2.01 | 84 | 171 | 60 | 110 | 164 | 105 | 106 |
| (2FB4) 22L-87L | 2.05 | 85 | 174 | 54 | 112 | 161 | 102 | 106 |
| Proteinase A | 2.03 | -87 | -70 | -64 | -146 | -108 | 100 | 104 |
| (2SGA) 42-58 | 1.90 | -93 | -76 | -58 | -145 | -107 | 107 | 100 |
| 191-220 | 2.05 | 101 | 68 | -76 | 95 | -55 | 104 | 105 |
|  | 2.16 | 93 | 63 | -82 | 96 | -44 | 106 | 105 |
| Wheat germ agglutinin | 2.00 | 90 | -62 | -59 | 144 | -60 | 105 | 109 |
| (3WGA) 17A-31A | 1.95 | 103 | -64 | -55 | 136 | -66 | 104 | 100 |

Table 3. The first column gives the protein names, the four letters code and the residue pairs for forming disulfide bond. The remaining columns are the following; d(S-S) the distance between the two sulfurs, $\chi$S-S the torsion angle $C\beta^i\text{-}S^i\text{-}S^j\text{-}C\beta^j$, $\chi^{i-1}$ and $\chi^{j-1}$ the torsion angles N-C$\alpha$-C$\beta$-S for residue i and j. $\chi^{i-2}$ and $\chi^{j-2}$ the torsion angles C$\alpha$-C$\beta$-S-S for residue i and j. $C\beta^i\text{-}S^i\text{-}S^j$ and $C\beta^j\text{-}S^j\text{-}S^i$ the bond angles. In each row the first line is the crystallographically observed data and the second line is the modeled the data. Note: a; In this case, variation of the dihedral angle $C\beta^i\text{-}S^i\text{-}S^j\text{-}C\beta^j$ is set by $|90| \pm 70°$.

### Example 5
### The MPMOD program

[0112] The present invention comprises a mini-protein modeling (MPMOD) program to perform a Monte Carlo search of conformational space. The idea for sampling Ramachandran maps was based on the program RANMOD (Sowdhamini et al., 1993). The inventors used Ponder and Richard's rotamer library (1989) along with a subroutine to generate side chains. Part of the program was written in standard Fortran-77 and the part of

the program that calculates the solvent accessible surface (SAS) energy and does the thermodynamic analysis was written in C (Hilser & Freire, 1996). FIG. 5 gives a flow chart of the program that is used to model disulfide bonds. In step 100, sequence, disulfide bond connectivity, and other parameters are inputted. The starting data is inputted manually or it may be retrieved from a database that is well known and used by those of skill in the art. From the input data, dihedral angles ($\phi$, $\psi$, $\omega$) are randomly generated in step 101. The generated dihedral angles are used to generate a polypeptide in step 102. Once the polypeptide is generated, a Van der Waals check is performed in step 103. If the van der Waals check is acceptable, then the SAS-based energy of the polypeptide is calculated in step 104. If the van der Waals check is not acceptable, then dihedral angles are regenerated. Next, step 105 searches for existing disulfide bonds in the generated polypeptide. After the disulfide bonds have been modeled, a van der Waals check is performed to ensure that the sulfer atom (S) is in good geometry with all the other atoms. The best position pairs are chosen and the SAS-based energy is calculated in step 106.

[0100] The MPMOD program is designed to generate disulfide bonded conformers or generate disulfide bonded conformers and linear conformers. If the program is run only to generate disulfide bonded conformers, then it is considered "the fast mod", which is illustrated in FIG. 13A and FIG. 13B. In step 200, sequence, disulfide bond connectivity, and other parameters are inputted. The starting data is inputted manually or it may be retrieved from a database that is well known and used by those of skill in the art. From the input data, dihedral angles ($\phi$, $\psi$, $\omega$) are randomly generated in step 201 and angles are assigned to each residue of the backbone. The generated dihedral angles are used to generate a backbone atoms, starting from three given atoms in step 202. The distance pairs are checked in step 203. It is important to determine the distance between the two C$\alpha$ atoms and the distance between the two C$\beta$ atoms. If the distance is not acceptable, then the dihedral angles are regenerated. The distance between the cysteines (C) plays a role in the rate of loop closure. If the distance is acceptable, then a Van der Waals check is performed in step 204. If the van der Waals check is acceptable, then the rest of the backbone is generated in step 205. If the van der Waals check is not acceptable, then dihedral angles are regenerated. While generating the backbone, if the van der waals check remains acceptable, then modeling of the disulfide bonds is performed in step 206. If the van der waals check does not remain acceptable, then dihedral angles are regenerated. Next, rotamers or side chains are added to the backbone in step 207. Rotamers are added to each residue except for the cysteines. From

30

step 207, one can collate all the none-van der waals violations in step 208 and regenerated dihedral angles and in step 209 the backbone and all rotamer combinations are written to a file. If the van der Waals check is acceptable for each rotamer in step 210, then disulfide bonded pairs are checked to ensure that the sulfer atom (S) is in good geometry with all the other atoms in step 211. If all the checks are acceptable, then the backbone angles and other information are written to a file in step 212. Next, a binding test is performed in step 213 for each conformer with the receptor to determine which conformer has a higher binding affinity. Finally, the SAS-based energy is calculated in step 214.

[0100] As mentioned the MPMOD program also can generate disulfide bonded conformers and linear conformers. This type of program is considered "the slow mod", which is illustrated in FIG. 14A and FIG. 14B. In step 300, sequence, disulfide bond connectivity, and other parameters are inputted. The starting data is inputted manually or it may be retrieved from a database that is well known and used by those of skill in the art. From the input data, dihedral angles ($\phi$, $\psi$, $\omega$) are randomly generated in step 301 and angles are assigned to each residue of the backbone. The generated dihedral angles are used to generate a backbone atoms, starting from three given atoms in step 302. Next, the rest of the backbone is generated in step 303. If the van der Waals check is acceptable, rotamers or side chains are added to the backbone in step 304. Rotamers are added to each residue. After the rotamers are added, the distance pairs are checked, modeling of the disulfide bonds and van der Waals check for the SS pairs with the complete conformer in step 305. If any step in 305 is unacceptable, the number of the conformer that can not form a SS bond is recorded and the program is linked to the COREX program to calculate the SAS-based energy $\Delta G$ for each conformer in step 308. If all steps in step 305 are acceptable, then the number of the conformer SS bond is recorded, the SAS-based energy $\Delta G$ for each conformer is calculated in step 306. After the calculations, each conformer is written to a file in step 307.

[0115] Yet further, the MPMOD program is capable of performing loop generation as shown in FIG. 15. In step 400, two residue numbers of the flexible loop of the protein and the accuracy are inputted. The starting data is inputted manually or it may be retrieved from a database that is well known and used by those of skill in the art. From the input data, dihedral angles ($\phi$, $\psi$, $\omega$) are randomly generated in step 401 and angles are assigned to each residue of the backbone. The generated dihedral angles are used to generate a backbone atoms or mainchain atoms. The distance pairs are checked in step 403. It is important to determine the distance between the two C$\alpha$ atoms and N and C terminals of the conformer.

31

In step 404, the distance between N and C terminal of the conformer is minimized by altering or modifying the dihedral angles. Step 405 requires that the handness of the conformer be same as the cutting parts of the target protein. Van der Waals check is performed in step 406 of the mainchain atom pairs. If the van der Waals check is acceptable, then the conformers are aligned to the target protein in step 407. Van der Waals check is performed in step 408 of the mainchain and target protein atom pairs. If it is acceptable, then rotamers or side chains are added to the mainchain in step 409. If the van der waals check is acceptable for each rotamer in step 409, then information is written to a file in step 410.

[0100] For the disulfide bond modeling module of the MPMOD program is illustrated in FIG. 16. In step 500, coordinates of N, Cα and Cβ of the two cysteines are obtained. Next, in step 501, a distance check is performed for Ca to Ca and Cb to Cb. If the distance is not acceptable, then other coordinates must be obtained in step 500. If the distance is acceptable, then the SG is generated on the circle formed by the rotation along Cα-Cβ bond in step 502. Next, bond length, bond angle, and dihedral angles are determined in step 503. If the measurements in step 503 are acceptable, the disulfide bond is formed and the coordinates are written to a file in step 504.

[0117] The binding test module of the MPMOD program is illustrated in FIG. 17. Step 600, requires pdb coordinates of the generated conformers and the crystal structure, segment of sequence for both alignments, criteria for best alignments, and three options for test "binding". Once all the information is gathered, the conformer is aligned to the corresponding peptide crystal structure in step 601. Next the root mean square deviation between each modeled conformer and the target peptide is determined and the average of conformational angle difference between each residue of the two conformers is determined. If the values are acceptable, then van der waals check of each conformer with the protein is performed in step 603. If the van der waals check is acceptable, then the SAS-base energy for each conformer is calculated in step 604 and the statistics are preformed in step 605.

## Example 6
### The loop closure rate for the cyclic peptides CXC and CXXC

[0118] Conformational ensembles for a series of CXC and CXXC, where X is one of the amino acids Ala, Val, Pro and Gly, were generated. Each ensemble consisted of 4000 conformers that can form a disulfide bond. The side chain was added to the backbone for each conformer. Only the backbone hydrogen atoms (HN and HCA) were generated in each conformer.

32

[0119] The disulfide bond loop closure "rate" or probability may be defined as $N_c/N_o$ where $N_c$ is the number of conformers that can potentially form disulfide bond and $N_o$ is the number of conformers that can not form a disulfide bond but have passed van der Waals check. This definition is similar to that of the equilibrium constant $K_c$. Table 4 listed the rates of loop closure from the modeling and experimental data for each type of peptide. The relative values from the modeling depict the same pattern as the experimental data. The rate for CPC is the highest in the CXC series and CPPC is has the lowest rate in the CXXC series. To directly compare both values, they were scaled to the same level. The common scale factor for CXC and CXXC can be defined by $K=\Sigma(Exp)/\Sigma(Mod)$, where $\Sigma(Exp)$ is the summation of all the experimental values, and $\Sigma(Mod)$ is the summation of all modeled values. Each individual value of the modeled conformer is multiplied by the scale factor K. FIG. 6 gives the comparison of scaled values of modeling with the experimental values. The values are in agreement for the CXXC series.

[0120] Table 4 shows that the rates of loop closure for the CXC series are much lower than those for CXXC series. This is determined by two factors, the distance between the two cysteines and the flexibility of the backbone. Statistics shows that to form disulfide bond, the distance $C_\alpha^i$-$C_\alpha^j$ between the two Cα atom of cysteine i and j must be within 4.0 to 7.0 Å and the distance $C_\beta^i$-$C_\beta^j$ between the two Cβ atoms must be within 3.3 to 4.7Å. The inventors surveyed all of the conformers that passed van der Waals check and found that for the CXC series the average distance for the two Cαs was about 6.2~6.5Å and for the two Cβs the distance was about 7.1~7.9Å. The distance of CPC was the shortest. The averaged Cβ distance of the randomly generated conformers is far from the suitable distance. Due to CXC only having three residues, the degree of flexibility of the backbone is not high enough to make the Cβ distance shorter unless the standard bond angles and bond lengths change. The ratio $N_{cacb}/N_{vdW}$, where $N_{cacb}$ is the number of conformers that have suitable Cα and Cβ distances and $N_{vdW}$ is the total number of conformer that passed van der Waals check, is 0.72%, 0.63%, 1.45% and 0.23% for CAC, CVC, CPC and CGC respectively. For the CXXC series, the average distance for the two Cαs and the Cβs are respectively 8.4~8.7Å and 9.1~9.6Å, with CPPC being the shortest. These distances are further away from the standard distances for forming disulfide bonds, but the residues have a much higher degree of flexibility for the backbone. A higher percentage of the conformers have suitable Cα and Cβ distances. The ratio $N_{cacb}/N_{vdW}$ gives 2.19%, 2.61%, 0.53% and 1.26% respectively for CAAC, CVVC, CPPC and CGGC.

33

[0121] To form a disulfide bond it is not just important to have a suitable $C\alpha$ and $C\beta$ distance, but also for the two sulfurs to have good steric positions. The inventors checked the number of conformers that satisfied the $C\alpha$ and $C\beta$ distances and the number of conformers that formed a disulfide bond. Although the conformers of CAC have the suitable $C\alpha$ and $C\beta$ distances, the probability of forming a disulfide bond is still smaller than CAAC because of CAC lacking a set of suitable geometrical parameters such as the bond length S-S, bond angle $C\beta$-S-S and the torsion angle $C\beta$-S-S-$C\beta$. The ratio $N_c/N_{cacb}$, where $N_c$ is the number of conformers that can form disulfide bonds, for the CAC, CVC, CPC and CGC series are 5.7%, 4.6%, 6.0% and 8.4% and the ratio $N_c/N_{cacb}$ for CAAC, CVVC, CPPC and CGGC are 36.0%, 35.1%, 34.3% and 33.0%. Therefore, the CXC series not only has a lower percentage of the conformers that have suitable $C\alpha$ and $C\beta$ distances, but also have a lower percentage of the conformers in which the two sulfurs to have good geometrical positions to form a disulfide bond. These factors led to a lower probability of loop closure for CXC than for CXXC.

**TABLE 4**

| Amino | CXC | | CXXC | |
|-------|-----|-----|------|-----|
| Acids | Mod | Exp | Mod | Exp |
| Ala | 0.091 | 3.1 | 0.604 | 71.0 |
| Val | 0.066 | 1.6 | 0.748 | 71.0 |
| Pro | 0.129 | 7.2 | 0.270 | 20.0 |
| Gly | 0.093 | 1.5 | 0.357 | 31.0 |

Table 4. The probability from modeling (labeled by Mod) is defined as $N_c/N_o$, where $N_c$ is the 4000 conformers that can former S-S bond and $N_o$ is the number of conformers that do not form SS bond but have passed vdW check. The equilibrium constant $K_c$ (labeled by Exp) is defined as $k_c/k_o$, where $k_c$ is the loop closing and $k_o$ the loop opening rate constant (Zhang and Snyder, 1989).

[0122] Comparing the probabilities of loop closure for the peptides having the same number of residues, the probability for CPC is the largest and for CVC is smallest for the CXC series. On the contrary, the probability for CPPC is the smallest and for CVVC the largest for the series CXXC. This was caused by a special property of Pro and Val. The ratios of $N_c/N_{cacb}$ for the two series are about the same, so only the ratios of $N_{cacb}/N_{vdW}$ are

34

significantly different. Due to the pyrrolidine ring, the proline tries to pull the two cysteines toward each other, so that the two cysteines of CPC have a better $C_\beta^i$-$C_\beta^j$ distance. It has a higher chance of forming a disulfide bond than other CXC series. In fact, the combination of XPX, where X is a non-proline amino acid, can easily form $\beta$ turn. On the other hand, Val has three rotamers $\chi^1$=180°, -60° and 60°. For the CVC peptide, the rotamer with $\chi^1$ =180 tries to push the two cysteines toward each other, but the rotamers, with $\chi^1$ = -60° and 60°, try to push the two cysteines away from each other, because the two CB branches are almost perpendicular to the backbone. It is more likely that the CB branch will push the two cysteines away from each other. This is why CVC has the smallest loop closure probability. As for the conformers of CXXC, the flexibility of the backbone plays a dominant role for the loop closure. Since the backbone of CPPC has a much lower flexibility than other peptides in the CXXC series, the chance of loop closure for CPPC is also lower than for other members of the CXXC series.

[0123] The inventors determined how many conformers were needed to get a meaningful ratio $N_c/N_o$. This ratio converges to a stable value as the number of conformers in the ensemble is increased. FIG.7 shows the ratio changes with increasing numbers of conformers for each series. When there are not enough conformers in the ensemble, the fluctuation of the values is large. As the number of conformers is increased, the ratio $N_c/N_o$ converges to a stable value. Therefore, the converged ratio may be compared with the experimentally measured result. From FIG.7 shows it is possible to over generate the number of conformers. One thousand conformers in each ensemble is enough conformers to get a converged ratio. The fluctuation after 1000 is not larger than 0.003% for the CXC series and not larger than 0.05% for the CXXC series.

## Example 7
### The longer $CX_nC$ polypeptide

[0124] Zhang and Snyder (1989) also measured the equilibrium constant $K_c$ for the series of $CA_nC$, where n is from 1 to 5. It was found that the $K_c$ constant decreases in the order of $CA_2C$, $CA_4C$, $CA_3C$, $CA_5C$, $CA_1C$, with an even numbers of $A_n$ high and odd numbers low (see line 3 of FIG. 8). The result of the inventors modeling using only the van der Waals approximation does not agree with Zhang and Snyder's (1989) experimental results. When the inventors increase the number of alanines between the two cysteines, the probability as defined by $N_c/N_o$ decreases monotonically after n>2 (see line 4 of FIG. 8). The

35

peak for $CA_4C$ was not captured by the modeling. This is not surprising because the inventor's calculation only considers geometrical factors, while the intramolecular interactions are more complicated in the experiment. The inventors performed a survey for the N-H...O hydrogen bond which was limited only to the backbone. The criteria for forming the H bond are $120° \leq \vartheta \leq 180°$ for the N-H...O bond angle and $d \leq 3.3 Å$ for the distance between N and O atoms. The ratio of numbers of hydrogen bonds in the disulfide bond closed conformers divided by the total number of conformer decrease in the order of $CA_2C$, $CA_4C$, $CA_5C$, $CA_3C$, $CA_1C$, is similar to that of $K_c$ constant (see line 1 of FIG.8 for ratio of H bonds). This indicates that the even numbered peptides $CA_nC$ are favored to have H bonds that stabilize the structure.

[0125] The solvent accessible surface (SAS) energy $\Delta G$ was calculated. Since the hydrogen bond is not considered in calculating $\Delta G$, compensation was given to the energy. For each H bond, the energy is increased 0.5 units. The energy weighted probability is defined as

$$\sum_{i=0}^{N_c} e^{-\Delta G_i / RT} \Bigg/ \sum_{i=0}^{N_a} e^{-\Delta G_i / RT},$$

which is the ratio of partition function for the closed peptides divided by the partition for all the unclosed peptides. This ratio follows the trend of experimental result (see line 2 of FIG.8). Some conformers in the ensemble that can form SS bond have a high probability, which leads to a high ratio. The peak of $CA_4C$ is slightly larger than $CA_3C$ and $CA_5C$.

## Example 8
### Construction of peptides for modeling the peptide-streptavidin complex

[0126] The backbone dihedral angles $(\phi, \psi)$ of the peptide were randomly generated in the four Ramachandran maps, one for glycine, one for proline, one for the CB-branched amino acids (VAL, ILE and THR), and one for all other amino acids. The trans and cis forms of the peptides have dihedral angles of $\omega = 180°$ and $0°$ with a small random deviation (usually within $\pm 5°$). The backbone of the peptide was generated based on the dihedral angles $(\phi, \psi, \omega)$ and the standard bond lengths and bond angles. Ponder and Richard's rotamer library (1989) was used to add side chains to the backbone. The simple hard sphere approximation was used to eliminate the grossly improbable conformers. Each atom was thought of as a hard sphere with its appropriate van der Waals radius. The minimum distances (Iijima *et al.*, 1987) between two atoms were used for the van der Waals check for each atom pair. These

36

distances are about 0.2 to 0.4Å shorter than the "normal" distance of Ramachandran *et al.*, (1963). If the backbone hydrogen atoms (HN and HCA) are generated, the overlap of the H atom with other atoms is even larger (about 0.5Å) than normal. Otherwise, it would not be efficient to generate the conformer due to van der Waals violations.

## Example 9
### Modeling disulfide bonds for the peptide-streptavidin complex

[0127] A single disulfide bond can be modeled using the method of the present invention. When two disulfide bonds are modeled in a conformer, attention should be paid to the computational efficiency. The probability to form two disulfide bonds simultaneously for a polypeptide is the product of the probabilities for each disulfide bond to form. Currently, it takes a long time to generate one peptide with two disulfide bonds. The inventors have created an efficient way to model a two-disulfide bond conformer. With two-disulfide bonded loops, the short one is modeled first. Conformations of this loop are fixed when the short one forms a disulfide bond. When the first loop is fixed, it may take long time to find the second loop if the first loop does not have suitable geometry. Therefore, some number of tries must be given to search for the second loop, while the conformation of the first loop is fixed. The number of tries is usually set to be between about 5 and 10. It is possible to obtain several polypeptides with one fixed conformation for the first loop and various conformations for the second loop. All conformers in the ensemble are kept for the "binding" test.

[0128] If the polypeptide is cycled by a covalent peptide bond (*i.e.*, the nitrogen (N) of the first residue makes a covalent bond with the carbon (C) of the last residue) the method for modeling the disulfide bond is no longer valid. The criteria to form such cyclic peptides are 1.35±0.6 Å for the N-C bond length and 120±35° for the bond angles (CA-N-C or CA-C-N). It is less efficient to generate such cyclic peptides than to generate a one disulfide-bonded peptide, since the former is searched only by one position of the atom N or C, whereas the latter is searched from a number of positions of sulfur.

## Example 10
### Aligning the conformers to the binding site of streptavidin

[0129] After ensembles of conformers were generated, the "binding" test was performed. The first step is to align the conformer to the template. The template is the peptide in the co-crystal structure complex. The second step is to screen the conformer by using the hard sphere potential model. For the peptide-streptavidin complex, the dominant binding force occurs at the HPQ sequence of the peptide, the modeled conformers were

37

aligned to the corresponding HPQ sequence of the crystal structure of the complex. Any higher resolution X-ray crystal structure can be used for the template. Two criteria were used to determine whether or not the alignment is successful. One criterion is the root mean square deviation (rmsd) between each modeled conformer k and the target peptide t.

$$\text{Rmsd}(k, t) = \left[\sum_{j=1}^{n}\left( (x(k, j)-x(t, j))^2 + (y(k, j)-y(t, j))^2 + (z(k, j)-z(t, j))^2 \right)\right]^{1/2}/n$$

where n is the number of atoms participating the alignment (n=9 for the HPQ sequence).

[0130] Another criterion is the average of conformational angle difference between each residue of the two conformers.

$$\Delta A(k, t) = \sum_{j=1}^{m} \left( |\phi(k, j) - \phi(t, j)| + |\psi(k,j) - \psi(t, j)| \right)/(2*m)$$

where m is number of residue for the compared sequence (m=3 for the HPQ sequence).

[0131] To determine whether the alignment is acceptable or not, two common reference values $\text{rmsd}_{ref}$ and $\Delta A_{ref}$, for rmsd(k, t) and $\Delta A(k, t)$ respectively, are given. For the kth conformers in the ensemble, if rmsd(k, t)$\leq$rmsd$_{ref}$ and $\Delta A(k, t)\leq\Delta A_{ref}$ are satisfied, then this alignment is acceptable. If any one of the criteria is not satisfied, the alignment is unacceptable and the conformer will be rejected. For the HPQ sequence, the reference values are $\text{rmsd}_{ref}$=0.50Å (Three atoms C$\alpha$, C and N were used for the alignment for each residue.) and $\Delta A_{ref}$=50°.

[0132] If the alignment is acceptable, a van der Waals check with streptavidin is performed as the second step to determine whether or not the final docking is successful. If there are any collisions for the atom pair of conformer and the target protein, the docking is not successful and the conformer is rejected. The atom radius for van der Waals check is the same as those mentioned before. If there is no van der Waals violation for any atom pair, the conformer is considered as being successfully docked into the protein. The "binding ratio" can be defined as the ratio $N_b/N_t$, where $N_b$ is the number of conformers that can be successfully docked into the HPQ binding pocket and $N_t$ is the total number of the conformers in the ensemble. The ratio correlated well with the experimentally measured binding affinity of the complex.

## Example 11
## Cluster analysis of the HPQ sequence

[0133] The inventors have surveyed the peptide-streptavidin complex. Table 5 lists the peptides and experimentally measured binding affinities with streptavidin. Ensembles of conformers for all these peptides have been generated following the above procedures. FIG. 10 gives an example of the ensemble for the peptide of CCHPQCGMVEEC. The HPQ sequence of the peptide is crucial for binding so it is necessary to know what fraction of the modeled conformers can adopt a type-I $\beta$ turn in the HPQ sequence. The crystal structure of CCHPQCGMVEEC (FIG. 9), determined at resolution 1.46Å, was used as the template to calculate that fraction. All the modeled conformers are aligned to the HPQ sequence of the crystal structure CCHPQCGMVEEC, using the reference values $rmsd_{ref}$=0.50Å and $\Delta A_{ref}$=50°. For each conformer, if the calculated $rmsd(k, t)$ and $\Delta A(k, t)$ are both less than the given reference values, the conformer is said to be "HPQ-like", or it is similar to the crystalstructure in the HPQ sequence. In other words, the modeled HPQ sequence can adopt a type-I $\beta$ turn. The percentage of conformers able to satisfy the criteria is listed in Table 5.

[0134] The HPQ-like conformer for the linear peptides (around 6%) is about 2-7 times smaller than the peptides with disulfide bond (12%-42%) (Table 5). The reason is that the linear peptides are not restrained in conformational space and can accept various conformations. Whereas, for the peptides with a disulfide bond, the configuration is constrained. The HPQ-like ratio for the linear peptides does not vary much. The ratio for the cyclic peptides varies according to the type and number of amino acids between the two cysteines. The only difference between the conformer AECHPQFNCIEGRK and AECHPQFPCIEGRK is at residue 8. But the ratios for both have a significant difference in which the former has a ratio of 22.4% and the latter has a ratio of 41.9%. Having a proline as residue 8 greatly increased the chance to form a type-I $\beta$ turn for the HPQ motif.

TABLE 5.

| Peptide | Observed | | Modeled | |
|---|---|---|---|---|
| | $K_D$ (µM) | $K_A$ (µM$^{-1}$) | HPQ-like (%) | Binder ($f_b$) (%) |
| **HDHPQNL**[1] | 282 | 0.004 | 4 | 1.3 |
| **FSHPQNT**[1] | 125 | 0.008 | 6 | 1.4 |
| **AHPQFPAEK**[4] | 136 | 0.007 | 6 | 1.50 |
| **AHPQFGAEK**[4] | 204 | 0.005 | 6 | 0.85 |
| **AECHPQGPPCIEGRK** [2] | 0.23 | 4.35 | 24.9 | 11.3 |
| **AECHPQFPCIEGRK**[2] | 0.93 | 1.08 | 41.9 | 28.7 |
| **AECHPQFNCIEGRK**[2] | 7 | 0.14 | 22.4 | 11.3 |
| **AECHPQFCIEGRK**[2] | 0.47 | 2.13 | 16.1 | 6.6 |
| cyclo-**CHPQFC**[2] | 0.27 | 3.70 | 18.7 | 16.9 |
| cyclo-**CHPQGPPC**[2] | 0.67 | 1.49 | 23.5 | 12. 0 |
| cyclo-**(AHPQFPAE)K**[4] | 0.13 | 7.69 | 21.4 | 21 |
| cyclo-**(AHPQFGAE)K**[4] | 19 | 0.05 | 12 | 7.1 |
| cyclo-**(AQYGHFAE)K**[4] | >5000 | 0.0002 | | |
| **RCCHPQCGMVEEC**[5] | 1.3 | 3.3 | 27.8 | 7.5 |
| **RCCHPQCGMAEEC**[5] | 2.3 | 0.45 | 25.3 | 7.2 |
| **RCCHPQFEPCMGC**[5] | 0.33 | 3.0 | 19.6 | 7.4 |

The list of experimentally observed binding and the modeled "bind ratio" fb.
1. Weber *et al.*, (1992) Biochemistry 31, 9350-9354.
2. Giebel *et al.*, (1995) Biochemistry 34, 15340-15435.
3. Schmidt *et al.*, (1996) J. Mol. Biol. 255, 735-766.
4. Zang *et al.*, (1998) Bioorganic & Medical Chemistry Letters 8, 2327-2332.

[0135] In the peptides, CCHPQCGMVEEC and CCHPQCGMAEEC the first two cysteines are too close to each other to form a disulfide bond. The combinations of the disulfide bonds that can be formed are the crossed form C1-C6, C2-C12, and the nested form C2-C6, C1-C12. The crossed form has a higher percentage of HPQ-like conformers than the nested form. This was caused by the smaller loop. Zhang and Snyder (1989) showed the equilibrium constant $K_c$ for forming CXXXC is smaller than for forming CXXXXC. The first loop in CCHPQC for the crossed form adopts higher ratio of type-I β turn in the HPQ sequence. When Ala is replaced by Val for peptides with two-loops, the fraction of HPQ-like conformers increase. The CB branched amino acid further limits the conformation of the HPQ sequence, which enhanced the ratio of HPQ-like conformers.

## Example 12
### The "binding ratio" of peptide-streptavidin complex

[0136] The X-ray co-crystal structure shows that all of the peptides bind to streptavidin at the same site. The HPQ sequence is crucial for the binding of the complex. When the HPQ motif of the modeled conformers is similar to that of the corresponding crystal structure, the modeled conformer has the potential to bind with streptavidin. Each HPQ-like conformer is aligned to the HPQ sequence of the co-crystal structure. If the conformer does not have a van der Waals collision with the target protein, it is defined as a "binder". The larger the fraction of "binder" in the ensemble, the higher the binding affinity is for the complex. The last column of Table 5 gives the percentage of "binder" in the ensembles. The fraction of "binder" correlates with the experimentally measured binding affinity for the series of peptides. The linear peptides are adopted by streptavidin at very low percentage (from 0.85% to 1.1%) compared with the cyclic or disulfide bonded peptides (from 7% to 28.7%). The measured binding affinity for the linear peptides is also much lower than the other peptides. This is caused by the entropy effect. The linear peptides are not constrained in conformational space and lose more entropy when they bind to the target protein. Therefore, the measured binding affinity and calculated "binder" fraction for the linear peptides is very low.

[0137] The last two peptides listed in Table 5 were selected from a phage display library. There are two disulfide bonds in each peptide. The conformation is more restricted than the peptides with one disulfide. It may be reasonable to expect an even higher affinity than the cyclic peptides because the conformation is more restricted by the two disulfide bonds. The measured binding affinity is actually less than that of some of the cyclic peptides.

41

The modeled fraction of "binder" also behaves like the measured affinity. This may be caused by the geometry of the binding site for this system. Although the peptide is more rigid and has a higher fraction of HPQ-like conformers, the chance to collide with streptavidin is higher because the miniprotein is too large to properly fit the environment at the binding site. The penalty from the collision is even greater than the advantage from rigidity of the peptides. The number of times that each residue collided with streptavidin was counted, assuming that each atom on the peptide collides with streptividin only one time. FIG. 11 shows the number of collisions for each residue for the two disulfide-bonded peptides. The second loop containing residues 7-11 (GMVEE) collides with streptavidin more often than other residues.

## Example 13
### The correlation of the measured binding affinity with the modeled "binding ratio"

[0138] The difference of Gibbs free energy for ligands to bind with a protein can be written as $\Delta G_m = -RT^*Ln(K_a)$, where $K_a$ is the measured association constant. R is a constant and T is the temperature. The measured free energy is assumed to have a linear relation with modeled "free energy". $\Delta G_f = m^*\Delta G_c + b$, where $\Delta G_c = -RT^*Ln[f_b/(1-f_b)]$, $f_b$ is the modeled "binding" fraction. The slope m and the intercept b can be determined by minimizing the summation of the difference Res $= \sum_{i=1}^{N}(\Delta G_f - \Delta G_m)^2$, where N is the total number of the peptide listed in Table 5. FIG. 12 shows the correlation of the "binding ratio" with the observed binding constant $K_a$. The straight line is fitted by minimizing the summation Res $= \sum_{i=1}^{N}(\Delta G_f - \Delta G_m)^2$.

## Example 14
### Using MPMOD to develop toxin-based inhibitors of viral entry

[0139] Compounds are being developed to inhibit attachment and/or replication of alphaviruses, flaviviruses and arenaviruses. These pathogenic RNA viruses are potential biological weapons and are of general medical concern. Mouse brain membrane receptor preparations are used to select Langat virus variants that do not bind. The E protein genes of these variants are sequenced to find mutated regions that identify nucleotides responsible for binding. The recombinant protein are expressed and subjected to X-ray crystallographic structure determination. The cell receptor is also identified at this time by screening a cDNA library for binding to the Langat E protein with binding detected by immunoreactivity. Candidate cDNAs will be screened further to identify open reading frames. The putative receptor will be expressed in Sf9 cells. The cell receptor's identity will be confirmed by the

42

ability of Langat to infect transfected cells. The domain of the Langat E protein containing the site of receptor interaction will be overexpressed to provide material for phage display screening. Phage display technology is used to identify toxin-based compounds that bind tightly to domain III and/or the E protein and interfere with attachment and subsequent viral entry into the cell. Determination of the structure of the cell receptor allows additional templates for phage display to be constructed. Identified compounds will be tested for anti-Langat activity in Vero cells, then in the mouse model by intraperitoneal and aerosol challenge.

[0140] The inventors have determined that spiperone, a dopamine D2 subtype receptor antagonist, competes with Japanese encephalitis virus for binding to mouse brain MRP (membrane receptor preparations). Toxin-based anti-viral compounds are being designed based on families of 10-45 residue disulfide-rich conformational constrained toxins including apamin, tertiapin, serafotoxin and conotoxins, and the human hormone endothelin. Constrained peptide loops and more rigid toxin-based molecules are being used because the structural restraints allow the reduction of conformational entropy loss upon binding and thus increase the affinity of binding, extend the compound's bioavailability by reducing its sensitivity to proteases in the serum and increase the specificity of interaction for a single target by eliminating conformations that might bind to human proteins. The optimization of toxin analog sequences can be rationally guide by an NMR solution struture determination to identify the changes in conformation and dynamics. Because phage display technology is used, once a sequence is identified as effective as an anti-viral compound, variants can be quickly optimized against related Langat E proteins and envelope proteins of similar viruses. While the rigid scaffold of the toxins is adapted by the inhibitors, the sequences identified differ greatly from that of the wild-type toxin, eliminating any intrinsic toxicity. The use of disulfide bridged loop peptides and structured toxin-based libraries restricts the conformational space sampled by each sequence.

[0141] A rational structure-based incremental approach is pursued in parallel with strict blind combinatorial methodology. Phage display libraries containing random octamer sequence constrained at their ends with a disulfide bond are prepared. Tight-binding loop peptides are synthesized and tested for inhibition of viral entry. The crystal structure of inhibitory loop peptides in complex with the E protein is determined. Using MPMOD, a compact folded structure is designed to stabilize the observed loop conformation. That

peptide is synthesized and tested for binding and inhibitory effects on viral infection. Binding interactions are optimized by use of a phage display library of related sequences.

[0142] Antiviral agents are screened *in vitro* a cell culture assay. Monkey kidney Vero cell cultures are pr-treated with different concentrations of the test agent before infection with various dilutions of Langat virus. After cultures are infected with Langat, cells are overlaid with agar containing the test agent at the same concentration. Cultures are incubated and subsequently stained to quantify virus plaque formation in agent-treated vs. mock-treated cultures. Any agent that reduces virus plaque formation by 90% or greater is studied further.

[0143] *In vivo* model studies utilize 4-week outbred NIH Swiss mice treated with the test agent one day before, at and on each of four day following Langat virus challenge. Different concentrations and routes (intraperitoneal and intranasal) of agent administration are examined with intraperitoneal virus challenge. Mean day of death of mice is compared with mock-treated mice and determined efficacy of the test agents. Any potential agent is tested further by its ability to protect against aerosol challenge.

44

## REFERENCES

[0144] All patents and publications mentioned in the specification are indicative of the level of those skilled in the art to which the invention pertains. All patents and publications are herein incorporated by reference to the same extent as if each individual publication was specifically and individually indicated to be incorporated by reference.

U.S. Patent 4,554,101
U.S. Patent 5,284,760
U.S. Patent 5,348,867
U.S. Patent 5,354,670
U.S. Patent 5,366,878
U.S. Patent 5,389,514
U.S. Patent 5,446,128
U.S. Patent 5,475,085
U.S. Patent 5,500,353
U.S. Patent 5,508,192
U.S. Patent 5,571,698
U.S. Patent 5,618,914
U.S. Patent 5,635,377
U.S. Patent 5,672,681
U.S. Patent 5,674,976
U.S. Patent 5,759,817
U.S. Patent 5,780,279
U.S. Patent 5,789,166
U.S. Patent 5,821,047
U.S. Patent 5,824,520
U.S. Patent 5,866,344
U.S. Patent 5,922,545
U.S. Patent 5,929,237

Benedetti, E., *et al.*, (1983) *Int. J. Pept. Protein Res.* 22 1-15

Chakrabarti, P. and Pal, D. (1998), *Protein Engineering*, Vol.11, 631-647

Gavezzotti, A. (1983) *J. Amer. Chem. Soc.* 105, 5220-5225.

Giebel, L.B. *et al.*, (1995) *Biochemistry 34*, 15430-15435.

Hilser VJ and Freire E (1996) *J Mol Biol.* 262(5):756-72.

Iijima, H., *et al.*, (1987) *Prot: Struct, Funct, and Genet* 2, 330-339.

Janin, J., *et al.*, (1978). *J. Mol. Biol.* Vol 125, 357-386.

Jeffreys, H. and Jeffreys, B. S. (1950) "Method of Mathematical Physics," p122. Cambridge Univ. Press, London and New York.

Johannesson *et al.*, 1999, *J. Med. Chem.* 42:601-608.

Johnson MS, *et al.*, (1993) *J Mol Biol.* 231(3):735-52.

Kyte, J. & Doolittle, R.F. (1982) *J Mol Biol.* 1982 157(1):105-32.

MacArthur MW and Thornton JM (1993) *Proteins* 17(3):232-51.

Momany, F. A., *et al.*, (1975) *J. Phys. Chem.* 79, 2361-2381.

Nemethy, G. and Scheraga, H.A. (1965) *Biopolymers*, 3,155

Ponder, J. W. & Richards, F. M. (1987), *J. Mol. Biol.* Vol 193, 775-791

Press, W.H. *et al.*, (1986) *"Numerical Recipes"*.

Ramachandran, G. N. & Sasisekharan, V. (1968), A*dv. Protein Chem.* 23, 283-437

Ramachandran, G. N., *et al.*, (1963). *J. Mol. Biol.* 7, 95.

Scheraga, H. A. (1992), *"Reviews in Computational Chemistry "*, Vol.3 by Kenny B. Lipkowitz and Donald B. Boyd, VCH Publishers, Inc. New York.

Schmidt, T. G. M., *et al.*, (1996) *J. Mol. Biol. 255, 735-766*

Sowdhamini, R., *et al.*, (1993) *Protein Engng,* 6 873-882.

Sowdhamini, R., *et al.*, (1989), *Protein Engng*, 3 95-103.

Vita *et al.*, 1998, *Biopolymers* 47:93-100.

Weber, P. C., *et al.*, (1992) *Biochemistry 31*, 9350-9354

Weisshoff *et al.*, 1999, *Eur. J. Biochem.* 259:776-788.

Wells TNC, *et al.*, (1996) *Methods.* 10(1):126-34.

Zang, X., *et al.*, (1998) *Bio-organic & Medical Chemistry Letters. 8*, 2327-2332.

Zhang, R. and Snyder, H. (1989) *J. Biol. Chem.* 264, 18472-18479.

[0145] One of skill in the art readily appreciates that the present invention is well adapted to carry out the objectives and obtain the ends and advantages mentioned as well as those inherent therein. Methods, procedures and techniques described herein are presently representative of the preferred embodiments and are intended to be exemplary and are not intended as limitations of the scope. Changes therein and other uses will occur to those skilled in the art which are encompassed within the spirit of the invention or defined by the scope of the pending claims.

## CLAIMS

We claim:

1. A computer-assisted method for use in modifying a protein, said method comprising the steps of:

    i.        generating random conformational angles from a set of protein sequence data;

    ii.   ` generating a protein backbone using the conformational angles;

    iii.     performing a van der Waals calculation of the protein backbone;

    iv.    calculating a solvent accessible surface based energy of conformers that are generated in steps i-iii;

    v.      modeling disulfide bonds in the protein backbone;

    vi.    performing a van der Waals calculation for the disulfide bonds;

    vii.   calculating a solvent accessible surface based energy of conformers that are generated in steps i-vi; and

    viii.   creating the modified protein with structural characteristics found in the above steps.

2. The method of claim 1, wherein the modified protein has increased stability.

3. The method of claim 1, further comprising determining coordinate pairs of the disulfide bonds.

4. The method of claim 1, wherein the conformational angles are $\phi, \psi$, or $\omega$.

5. The method of claim 1, further comprising determining the number of conformers that are able to form disulfide bonds.

6. The method of claim 1, further comprising adding rotamers to each residue in the protein backbone.

7. The method of claim 1, further comprising performing a binding test for each conformer with a template molecule.

8. The method of claim 1, further comprising calculating the rate of disulfide bond loop closure.

9. The method of claim 8, wherein determining the rate of disulfide bond loop closure in the protein comprises the steps of

    i.        performing a van der Waals calculation on a multiplicity of conformers of the protein and subtracting those conformers that can not form an intramolecular disulfide to yield an ensemble of $N_0$ sterically allowed conformers;

ii.     analyzing the ensemble of sterically allowed conformers to yield an ensemble of $N_c$ conformers that are capable of forming an intramolecular disulfide bond; and

iii.    calculating the ratio $N_c/N_o$ which represents the rate of disulfide bond loop closure in the peptide.

10. A method of protein miniaturization comprising modeling a protein to have the necessary active site conformation using the method of claim 1 while reducing the total number of amino acids in the protein.

11. A method of increasing binding affinity between a protein and a template molecule by decreasing the conformational entropy loss upon binding by the protein comprising the constraint of at least one loop of an unstable region of the protein in conformational space using the method of claim 1.

12. A computer-readable storage medium having stored therein a software program which executes the steps of claim 1.

13. A computer-readable storage medium having stored therein a software program which executes the steps of claim 9.

14. A modified-protein with increased binding affinity produced by the method comprising the steps of:

i.     performing the steps of the method of claim 1; and

ii.    performing a binding test for each conformer with a template molecule; and

iii.   creating the modified protein using structural characteristics found in the above steps to increase binding affinity in the modified protein.

15. A modified-protein produced by the method comprising the steps of:

i.     performing the steps of the method of claim 1; and

ii.    creating the modified protein using structural characteristics found in the above steps to increase stability of the modified protein.

16. A computer-assisted method for use in modifying a protein comprising the steps of:

i.     generating random conformational angles in allowed region of Ramachandran maps from a set of protein sequence data;

ii.    generating a protein backbone using said conformational angles;

48

iii.    performing van der Waals calculation for each backbone atom;

iv.    determining disulfide bonds in the protein backbone;

v.    adding a rotamer to residues in the backbone;

vi.    performing van der Waals calculation for rotamers;

vii.    performing binding test with a template protein;

viii.    calculating solvent accessible surface based energy of all conformers that are generated in steps i-vii; and

ix    creating the modified protein using structural characteristics identified in the above steps.

17. The method of claim 16, wherein generating the protein backbone comprises assigning conformation angles to each residue of the backbone.

18. The method of claim 16, further comprising generating distance pairs between atoms.

19. The method of claim 16, further comprising determining coordinate pairs of the disulfide bonds.

20. The method of claim 16, wherein the backbone atoms are N, CA, or C.

21. The method of claim 16, wherein the conformational angles are $\phi, \psi$, or $\omega$.

22. A computer-readable storage medium having stored therein a software program which executes the steps of claim 16.

23. A computer-assisted method for use in modifying a protein comprising the steps of:

i.    generating random conformational angles in allowed region of Ramachandran maps from a set of protein sequence data;

ii.    generating a protein backbone using said conformational angles;

iii.    adding a rotamer to residues in the backbone;

iv.    determining disulfide bonds in the protein backbone;

v.    linking the method to a computer assisted program that calculates linear conformers;

v.      calculating solvent accessible surface based energy of conformers that are generated in steps i-iv; and

vi.     creating the modified protein using structural characteristics identified in the above steps.

24. The method of claim 23, wherein the computer assisted program that calculates linear conformers is COREX.

25. The method of claim 23, wherein the conformational angles are $\phi, \psi$, or $\omega$.

26. A computer-readable storage medium having stored therein a software program which executes the steps of claim 23.

27. A method for determining the rate of disulfide bond loop closure in a protein comprising at least one two-cysteine motif represented by $C\text{-}X_n\text{-}C$ where n is an integer, the method comprising the steps of:

i.      performing a van der Waals calculation on a multiplicity of conformers of the protein and subtracting those conformers that can not form an intramolecular disulfide to yield an ensemble of $N_o$ sterically allowed conformers;

ii.     analyzing the ensemble of sterically allowed conformers to yield an ensemble of $N_c$ conformers that are capable of forming an intramolecular disulfide bond; and

iii.    calculating the ratio $N_c/N_o$ which represents the rate of disulfide bond loop closure in the peptide.

28. The method of claim 27 wherein the rate is compared to the rate of disulfide-bond loop closure of the protein containing at least one different two-cysteine motif.

29. The method of claim 27 further comprising the step of generating peptide backbone coordinates for the $C\text{-}X_n\text{-}C$ motif from standard bond angles, bond lengths and dihedral angles randomly generated within the allowed regions of a Ramachandran map for each residue to yield the multiplicity of conformers of the protein.

30. The method of claim 29 further comprising the step of using a side chain rotamer library to generate $C\text{-}X_n\text{-}C$ side chain coordinates to yield the multiplicity of conformers of the peptide.

31. The method of claim 27 wherein analyzing the sterically allowed conformers comprises calculating the free energy of the conformers based upon the solvent accessible surface area.

32. The method of claim 31 wherein analyzing the sterically allowed conformers further comprises flexibly modeling the cysteine side chains.

33. The method of claim 27 further comprising the step of weighting $N_c$ and $N_o$ by the difference in free energy ($\Delta G$) between the dithiol and disulfide forms of the $C-X_n-C$ motif and calculating the ratio

$$\sum_{i=0}^{N_c} e^{-\Delta G_i / RT} \bigg/ \sum_{i=0}^{N_o} e^{-\Delta G_i / RT}$$

which represents the energy-weighted rate of disulfide loop closure in the protein.

34. The method of claim 33 further comprising the step of identifying an ensemble of $N_c$ conformers of the protein that can potentially form an intramolecular disulfide bond.

35. The method of claim 33 wherein docking the ensemble of $N_c$ conformers to a binding site on a template biomolecule comprises the steps of:

    i.      aligning the ensemble of $N_c$ conformers to a binding site on a template biomolecule to yield an ensemble of aligned conformers; and

    ii.     performing a van der Waals calculation on the ensemble of aligned conformers to yield an ensemble of $N_b$ sterically allowed conformers that bind to the template biomolecule.

36. The method of claim 35 wherein docking the ensemble of $N_c$ conformers to a binding site on a template biomolecule comprises the steps of:

    i.      aligning the ensemble of $N_c$ conformers to a binding site on a template biomolecule to yield an ensemble of aligned conformers; and

    ii.     performing a van der Waals calculation on the ensemble of aligned conformers to yield an ensemble of $N_b$ sterically allowed conformers that bind to the template biomolecule.

51

37. The method of claim 27 wherein the protein further comprises a plurality of two-cysteine motifs represented by $C-X_n-C$ wherein n is independently an integer for each two-cysteine motif.

38. A computer-readable storage medium having stored therein a software program which executes the steps of claim 27.

39. A method for assessing the binding affinity of a protein to a template molecule, wherein the protein comprises at least one two-cysteine motif represented by $C-X_n-C$ where n is an integer, the method comprising:

    i.      docking the ensemble of $N_c$ conformers to a binding site on a template biomolecule to yield an ensemble of $N_b$ conformers that bind the template biomolecule; and

    ii.      calculating the ratio $N_b/N_c$ which is indicative of the binding affinity of the protein for the template biomolecule.

40. A method for assessing the binding affinity of a protein to a template molecule, wherein the protein comprises at least one two-cysteine motif represented by $C-X_n-C$ where n is an integer, the method comprising the steps of:

    i.      screening a population of candidate peptides comprising at least one two-cysteine motif represented by $C-X_n-C$ where n is an integer to yield a plurality of candidate peptides that are capable of forming an intramolecular disulfide bond; and

    ii.      performing the method of claim 39 on at least one candidate peptide that are capable of forming an intramolecular disulfide bond to assess the binding affinity of the candidate peptide.

41. The method of claim 40 wherein the each candidate peptide comprises a preselected amino acid sequence.

42. The method of claim 41 wherein the preselected amino acid sequence predisposes the peptide to form a desired secondary structure.

43. The method of claims 42 wherein the desired secondary structure is a β-turn.

44. A method for modifying a protein comprising the steps of:

    i.      evaluating an X-ray crystal structure or a nuclear magnetic resonance solution structure comprising an oxidized reference peptide bound to a target molecule, the reference protein comprising at least one intramolecular disulfide bond, to identify at least two amino acids at positions favorable to intramolecular disulfide bond formation;

52

ii.     substituting cysteines for the two amino acids in the reference protein to yield a modified protein comprising at least four cysteines;

iii.    identifying an ensemble of $N_c$ conformers of the modified protein that can potentially form at least two intramolecular disulfide bonds;

iv.     docking the ensemble of $N_c$ conformers to the binding site on the template biomolecule to yield an ensemble of $N_b$ conformers that bind the template biomolecule;

v.      calculating the ratio $N_b/N_c$ which is indicative of the binding affinity of the modified protein for the template biomolecule; and

vi.     repeating steps (i.)-(v.) to yield modified proteins having cysteine substitutions at different positions and identifying modified peptides with the highest $N_b/N_c$ ratios.

45. The method of claim 44 wherein the identifying an ensemble step comprises the steps of:

i.      identifying a first conformer of the protein that are capable of forming a first intramolecular disulfide bond defining a first disulfide-bonded loop;

ii.     constraining the model by the first disulfide bond; and

iii.    identifying a second conformer of the protein that are capable of forming a second intramolecular disulfide bond defining a second longer disulfide-bonded loop.

46. The method of claim 44 wherein a second conformer is not identified after about 5 to about 10 attempts to identify said conformer, the method further comprising the steps of:

i.      eliminating the first disulfide bond from the model;

ii.     identifying a first conformer of the peptide that can potentially form a first intramolecular disulfide bond defining a different first disulfide-bonded loop;

iii.    constraining the model by the first disulfide bond; and

v.      identifying a second conformer of the peptide that can potentially form a second intramolecular disulfide bond defining a second longer disulfide-bonded loop.

47. A method for assessing the binding affinity of a protein to a template molecule, wherein the protein comprises a flexible loop, the method comprising the steps of:

    i.        generating a peptide conformation of length N from a starting residue I and matching to a target residue I + N on the peptide model;

    ii.       accepting the loop conformation when the deviation between residue N and the target residue is small;

    iii.      closing the loop using a geometric minimization method;

    iv.      selecting the residue conformation by the method of claim 29;

    v.        generating an ensemble of surface loops; and

    vi.      estimating the binding affinity by testing the docking of the full mini-protein ensemble and peptide target containing the loop ensemble.

48. A protein produced by protein miniaturization comprising modeling a protein to have the necessary active site conformation using the method of claim 1 while reducing the total number of amino acids in the protein.

49. A protein capable of docking into a binding site wherein the conformation of a portion of said protein was constrained by the introduction of a disulfide bond by the method of claim 1.

50. A protein, created by the method of claim 1, having the characteristic of inhibiting the binding of a virus to a cell wherein the protein is based upon a tertiary structure of a toxin and comprises at least one loop constrained by a disulfide.

51. An ensemble of intramolecular disulfide bond-forming conformers of said loop from the protein of claim 50.

52. A protein having decreased conformational entropy loss upon binding to a template molecule in comparison to the naturally occurring protein due to the constraint of at least one loop of an unstable region of a protein in conformational space by the formation of a disulfide bond other than disulfide bonds found in the naturally occurring protein using the method of claim 1.

53. An ensemble of intramolecular disulfide bond-forming conformers of said loop of the protein of claim 52.

54. A protein modified by the method of claim 44.

55. A computer system for designing a modified-protein, said system comprising:

ix.    a database containing a set of protein sequence data;

x.    a software program coupled with said database, the software program adapted for performing the steps of :

(a) generating random conformational angles from the set of protein sequence data,

(b) generating a protein backbone using the conformational angles,

(c) performing a van der Waals calculation of the protein backbone,

(d) calculating a solvent accessible surface based energy of conformers that are generated in steps (a) – (c),

(e) modeling disulfide bonds in the protein backbone;

(f) performing a van der Waals calculation for the disulfide bonds;

(g) calculating a solvent accessible surface based energy of conformers that are generated in steps (a) – (f) ; and

(h) creating the modified protein with structural characteristics found in the above steps.

56. A computer system for designing a modified-protein, said system comprising:

i.    a database containing a set of protein sequence data;

ii.    a software program coupled with said database, the software program adapted for performing the steps of :

(a)    generating randomly conformational angles in allowed region of Ramachandran maps from the  set of protein sequence data;

(b)    generating a protein backbone using said conformational angles;

(c)    adding a rotamer to residues in the backbone;

(d)    determining disulfide bonds in the protein backbone;

(e)    calculating linear conformers;

(f)    calculating solvent accessible surface based energy of conformers that are generated in steps  (a) – (d); and

(g)    creating the modified protein using structural characteristics identified in the above steps.

57. The computer system of claim 56, wherein the calculating step includes linking to an external program for calculating conformers.

1/21
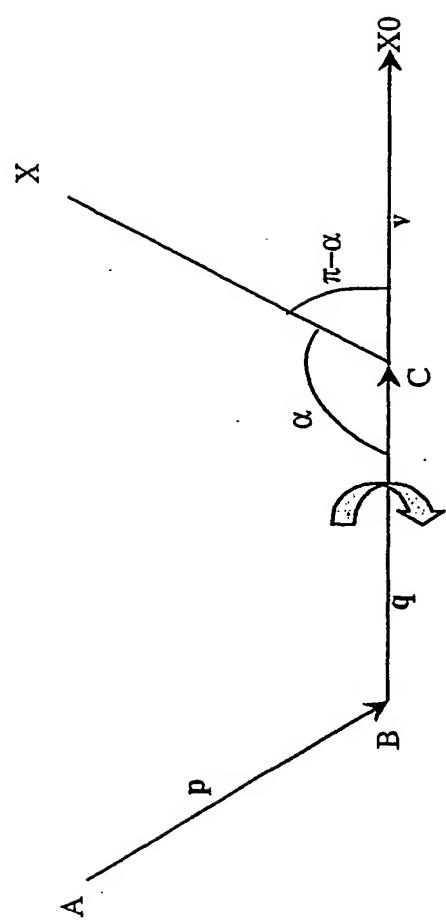


FIG. 1

2/21
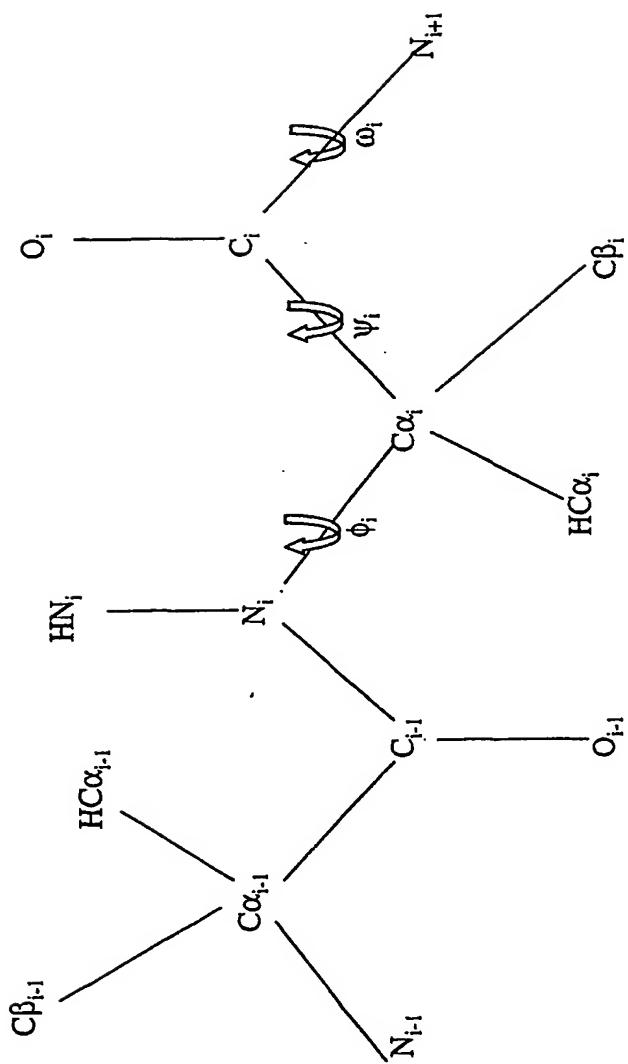


FIG. 2

3/21



FIG. 3

4/21



FIG. 3

5/21



FIG. 4

100 — Input: sequence, disulfide bond connectivity, and other parameters

101 — Generate φ, ψ, ω (randomly)

102 — Generate polypeptide

103 — Van der Waals check — *Bad*

*Ok*

104 — Record number of conformers and the coordinate.
Calculate SAS-based energy

105 — Disulfide bond modeling — *No*

*Ok*

106 — Record number of conformers and the coordinate.
Calculate SAS-based energy

FIG. 5

FIG. 6A

FIG. 6B

9/21



FIG. 7

FIG. 8

The peptide RCCHPQCGMVEECR
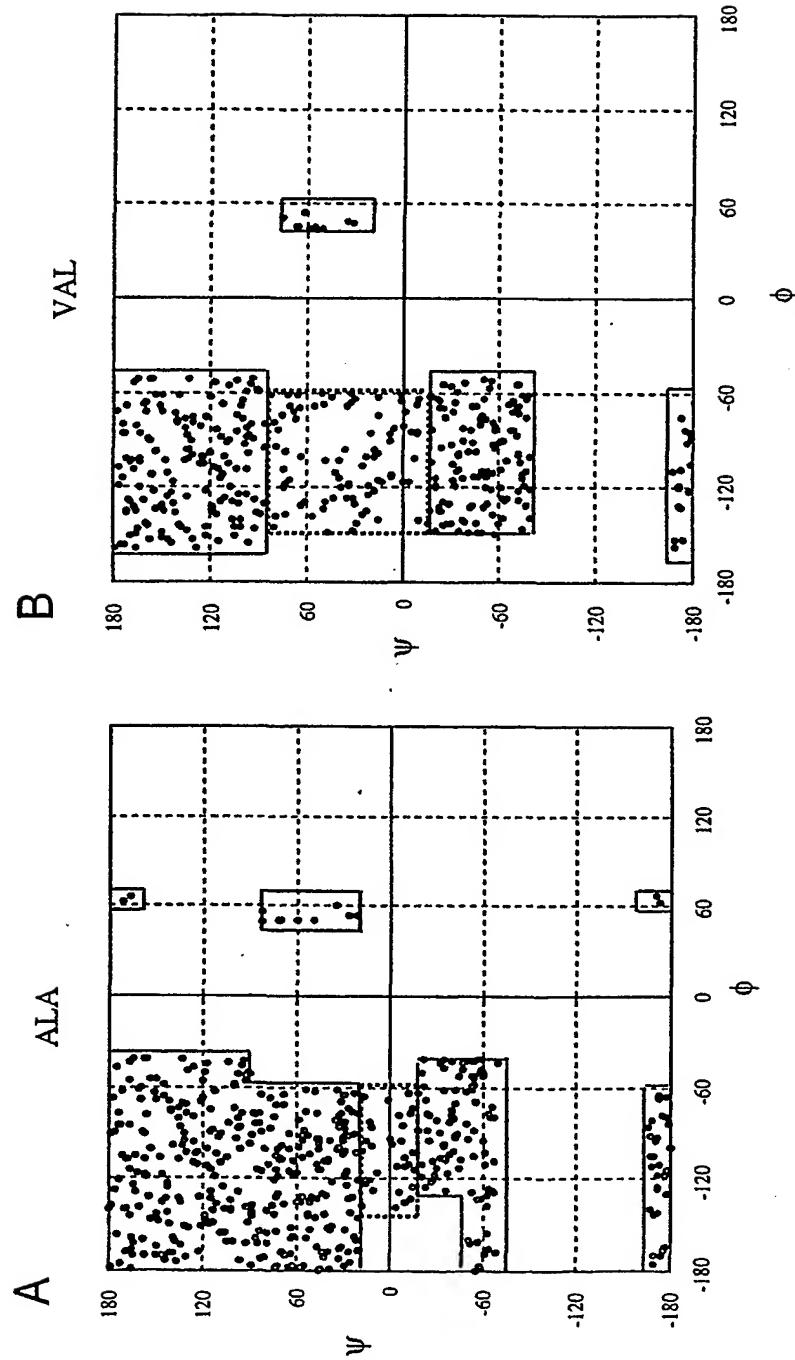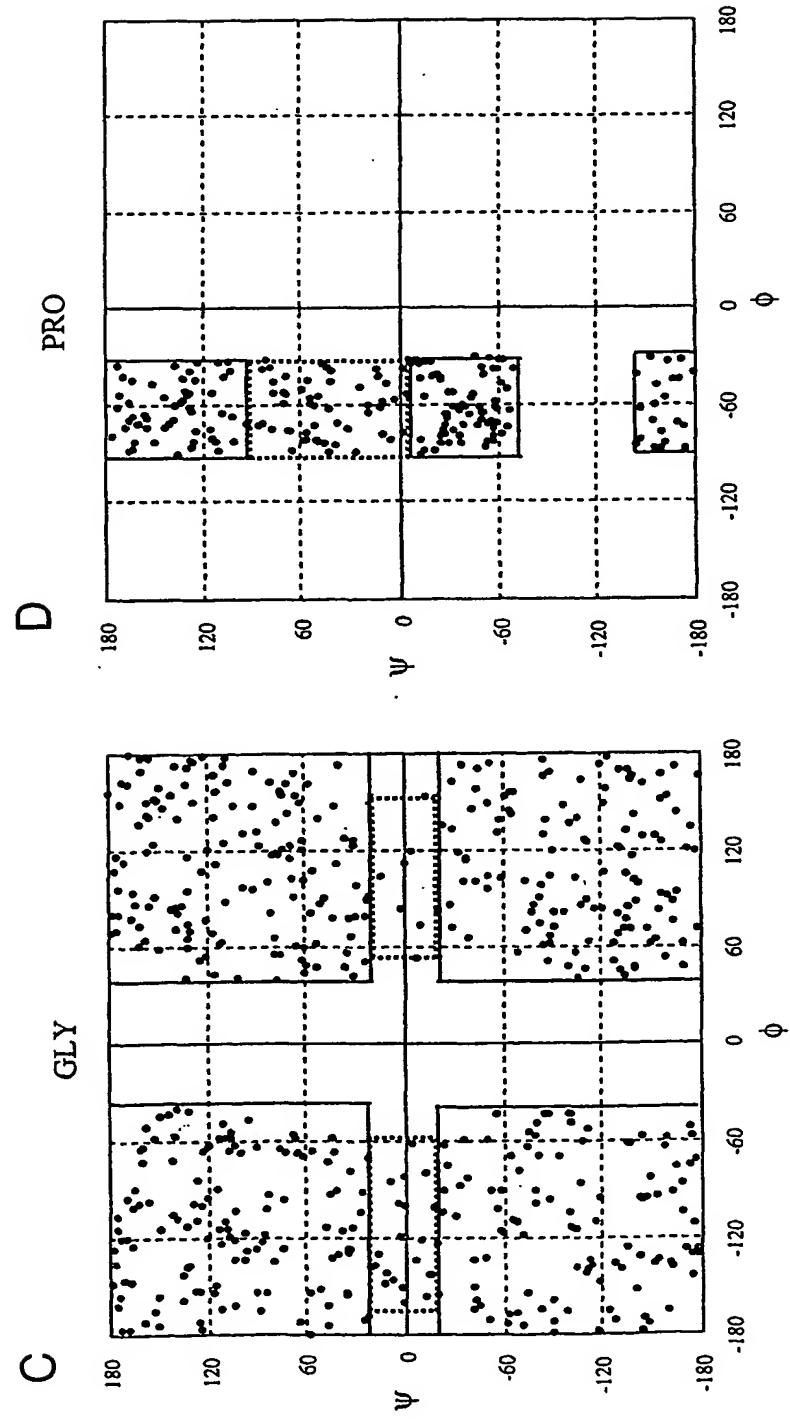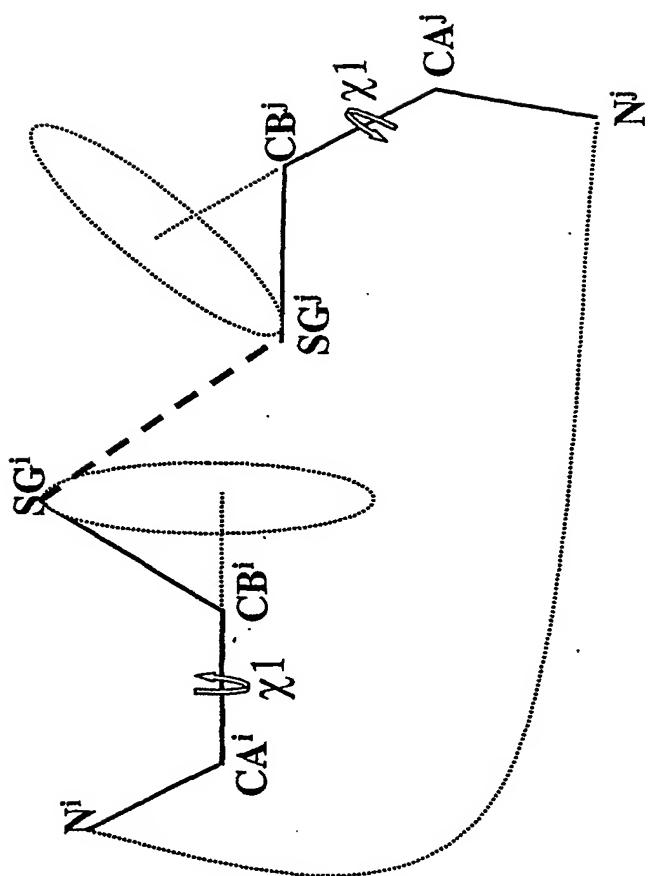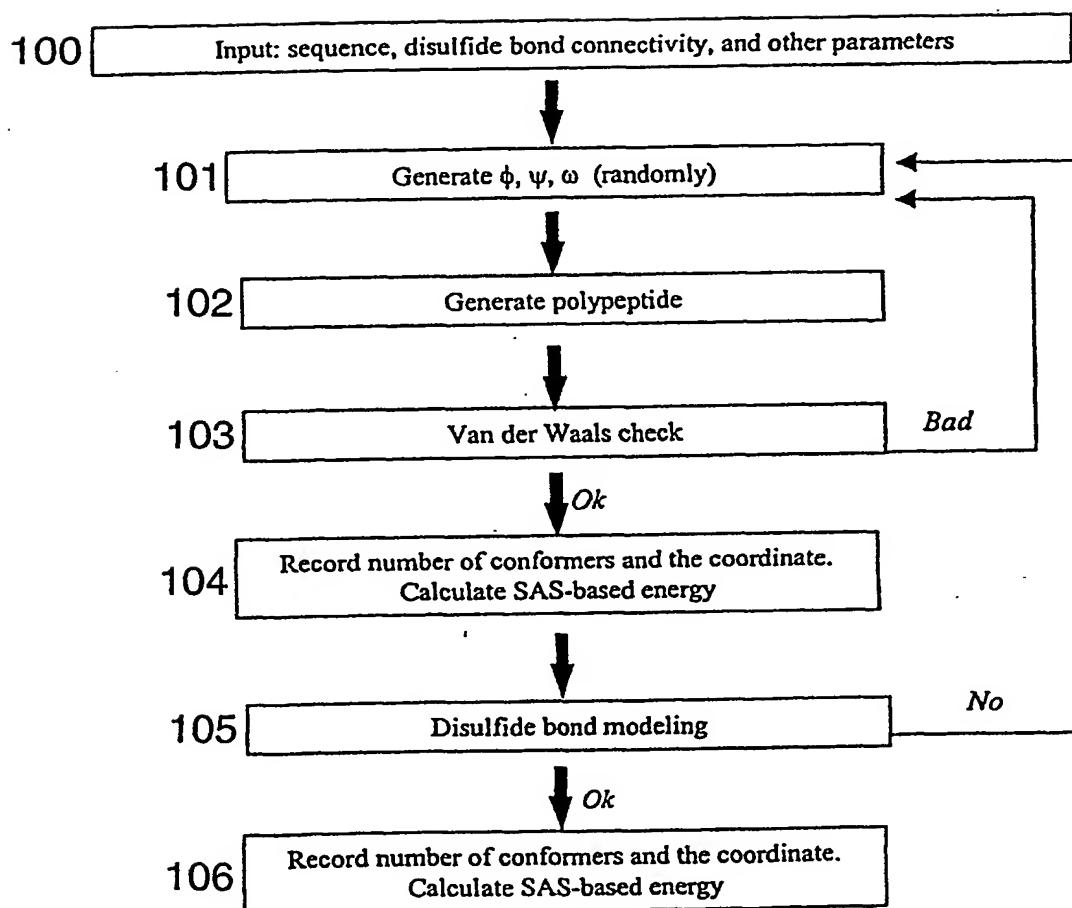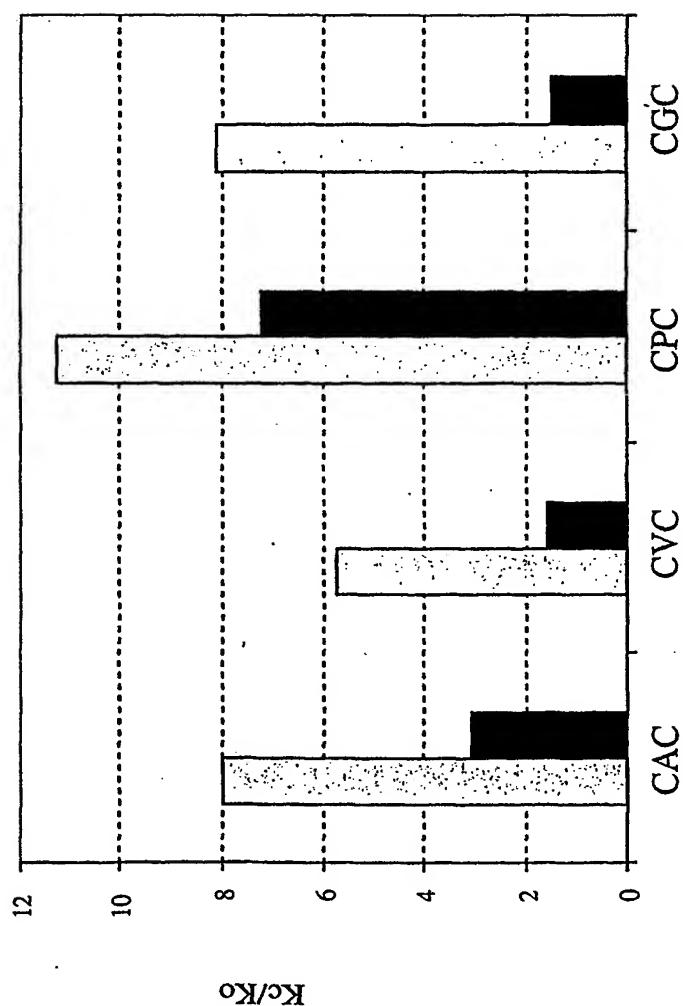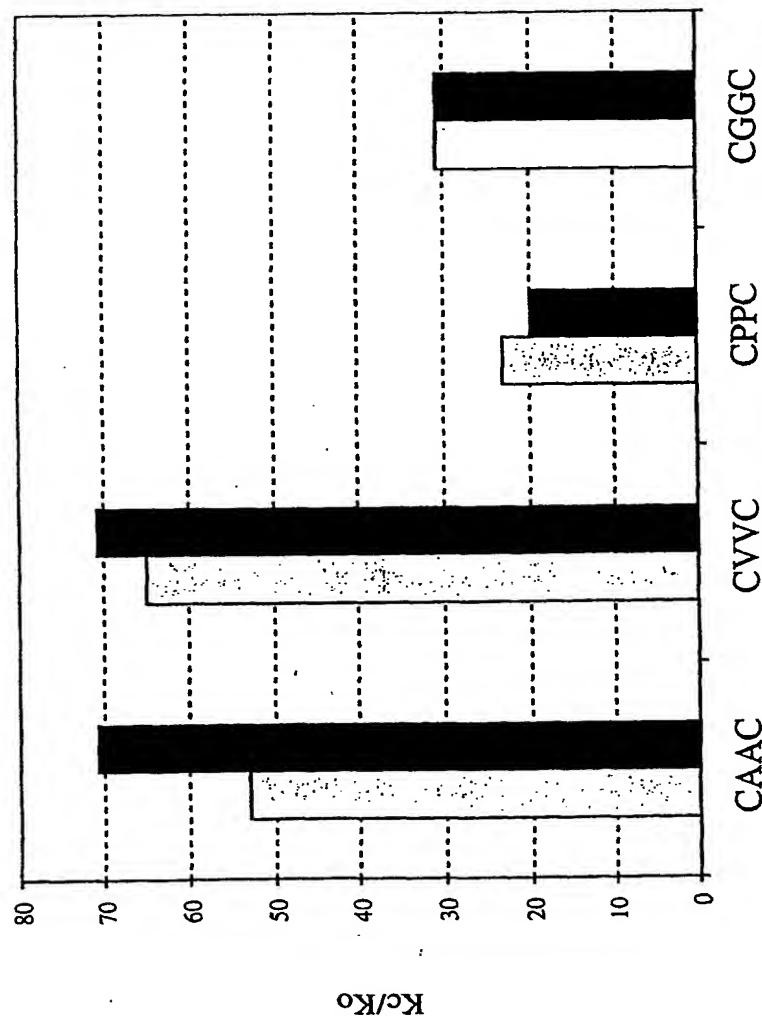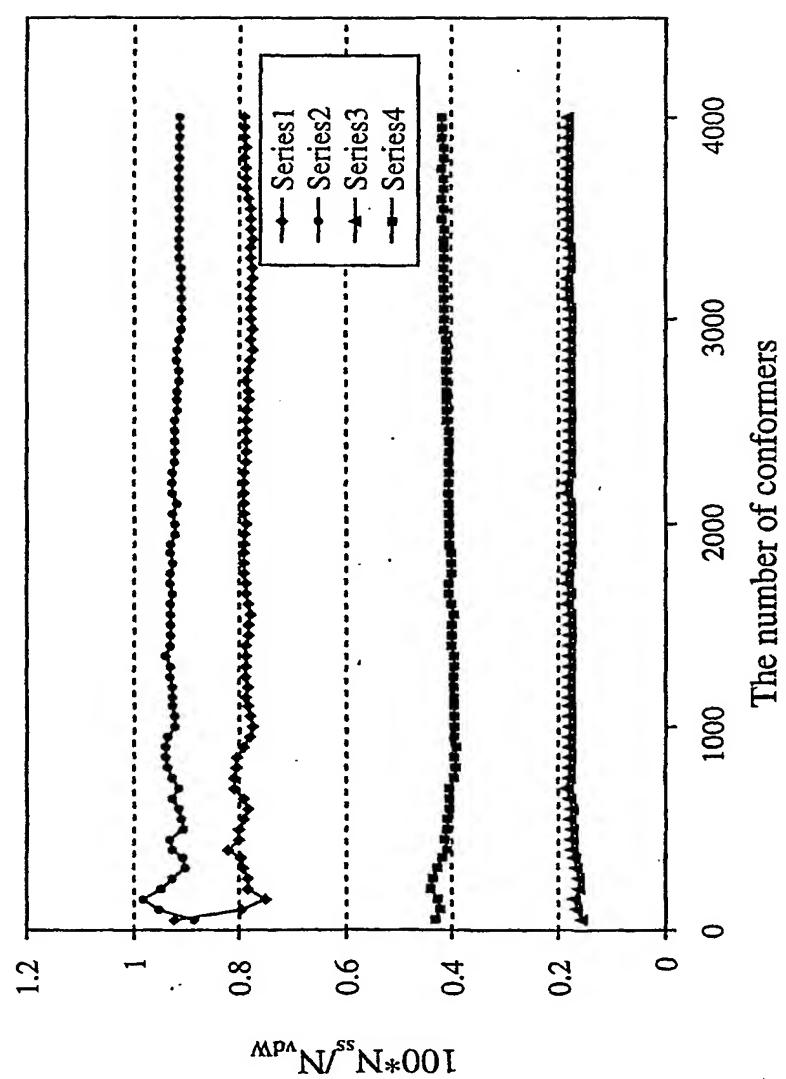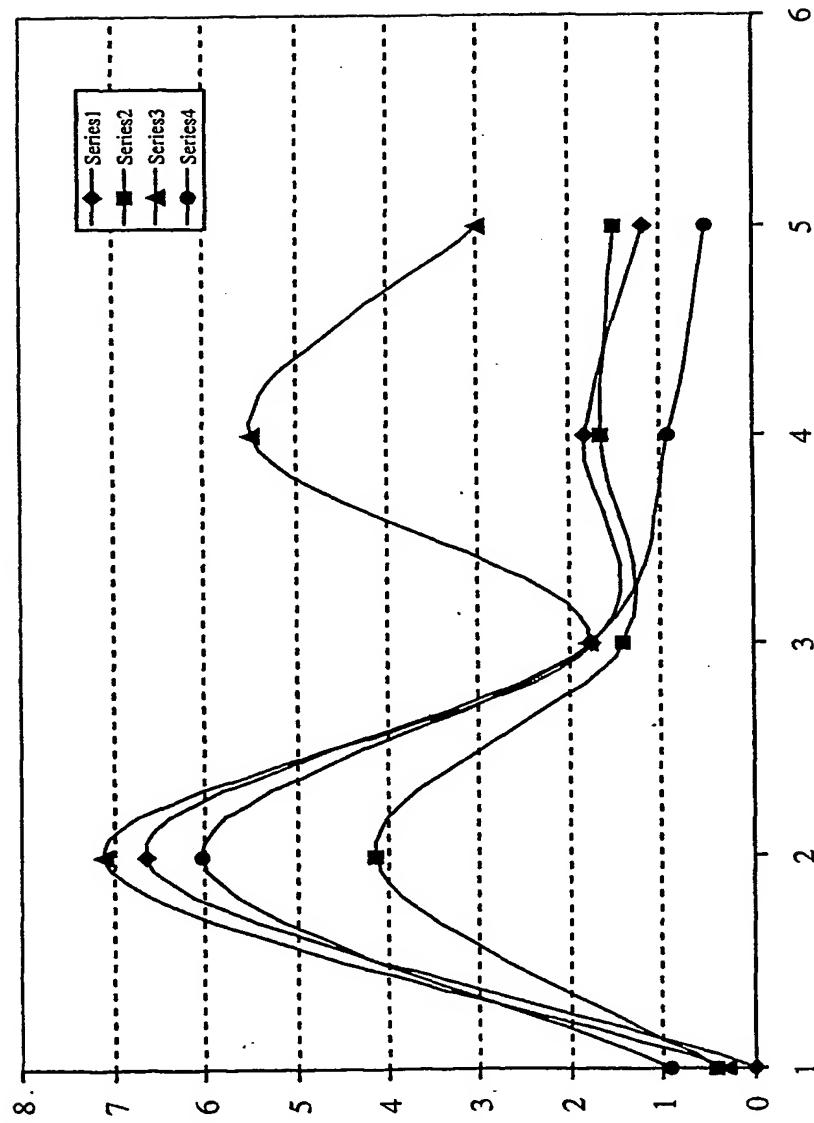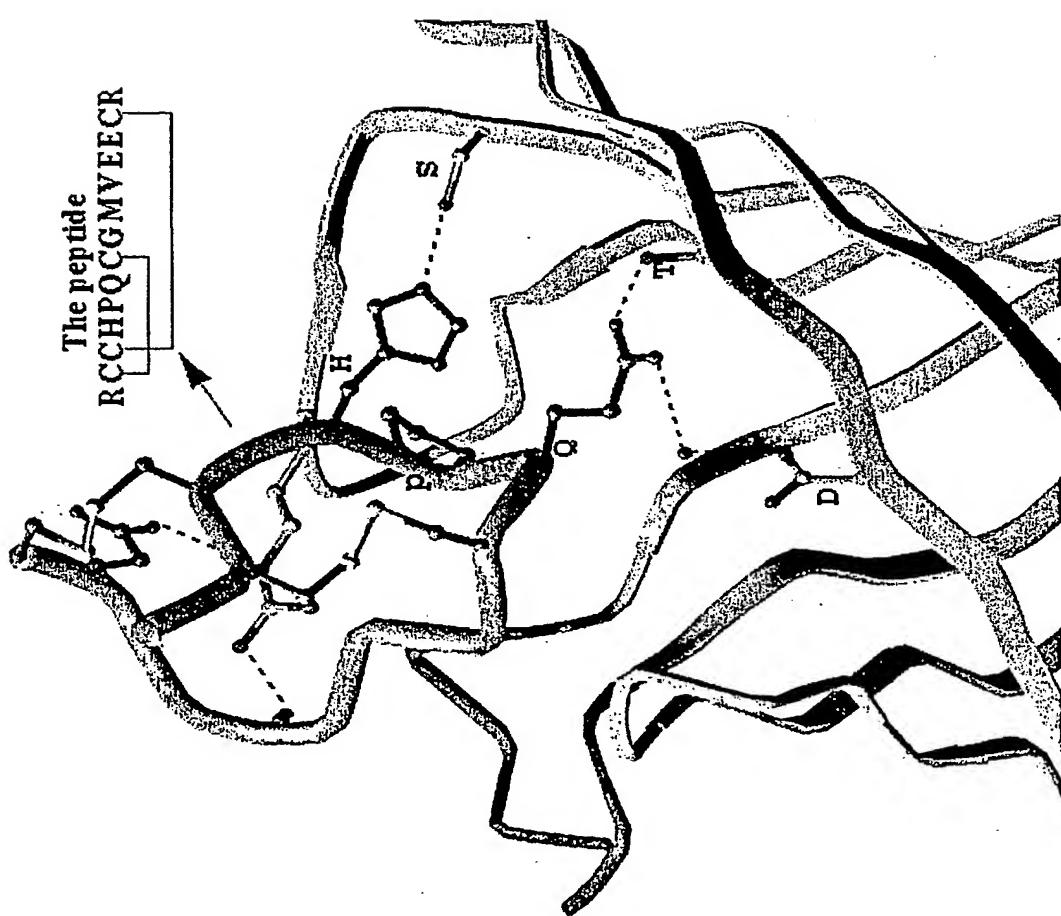
FIG. 9

12/21
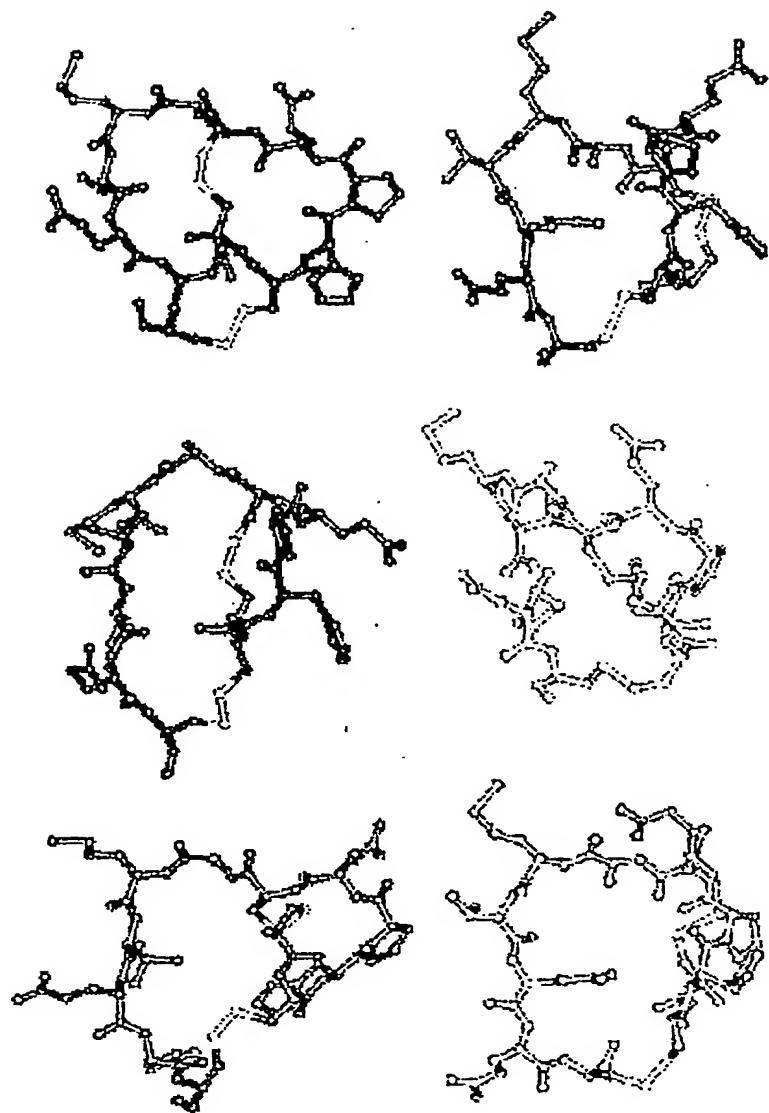


FIG. 10

13/21
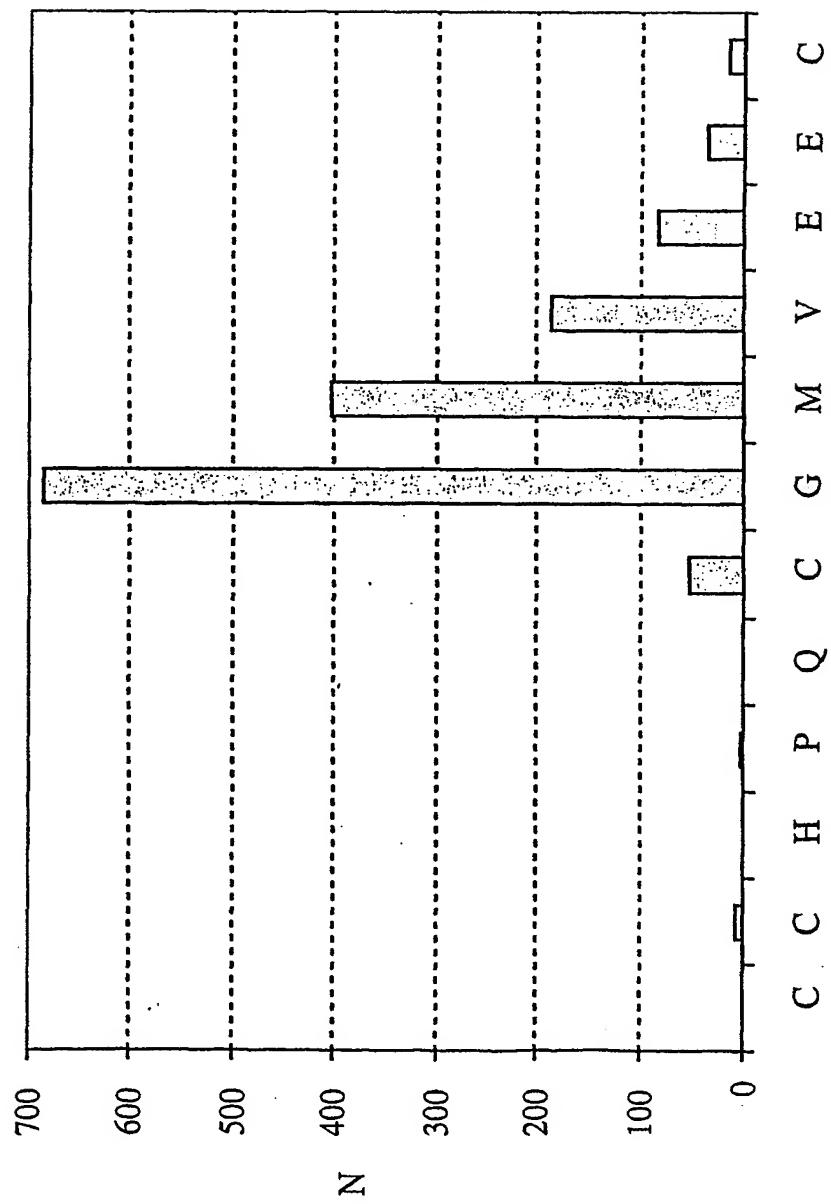


FIG. 11

14/21
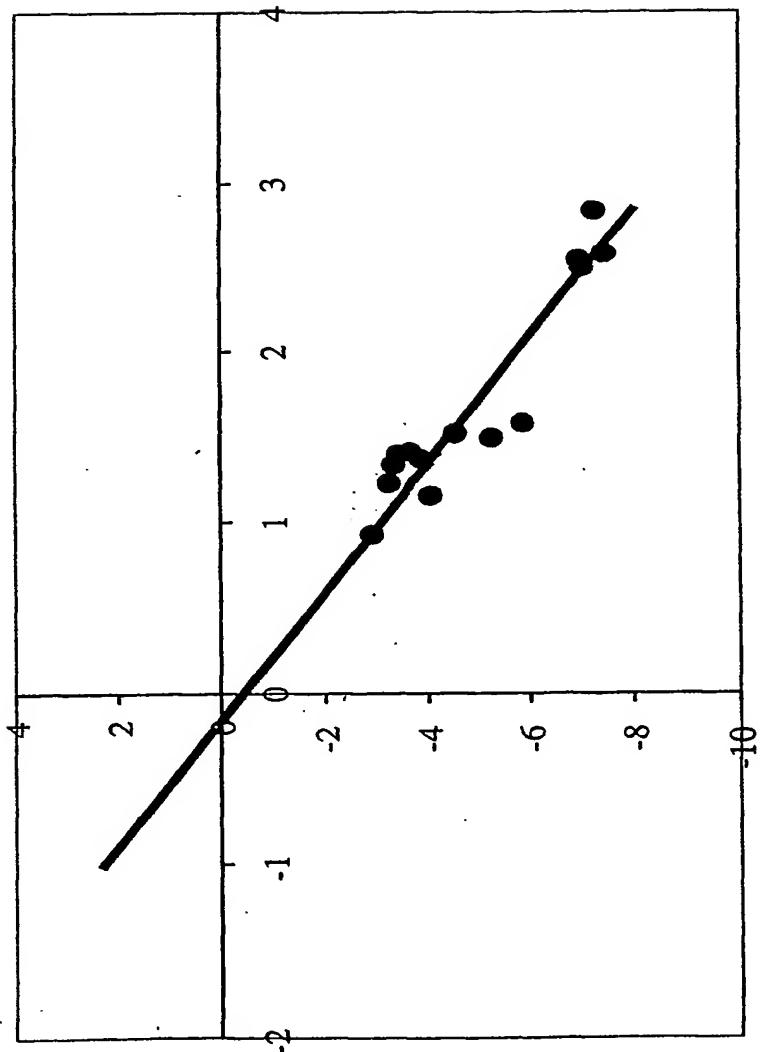


FIG. 12

$\Delta G_c = -RT * \ln[f_b/(1-f_b)]$

$\Delta G_m = -RT * \ln(K_a)$

15/21

# A flow chart of the program MPMOD
# (the fast mod)

200 | Input: peptide sequence, disulfide bond connectivity and other parameters

201 | Randomly generate Φ, Ψ, ω in the allowed regions of the Ramachandran maps (Gly, Pro, CB-branched, and the rest of the amino acid). Assign the angles to each residue of the backbone.

202 | Generate backbone atoms N, CA, C starting from three given atoms.

203 | Check the CA-CA and CB-CB distance pairs. CB were added to the CYS and deleted after using it.   → *Bad*

204 | Do vdw check for the backbone atoms (N, CA, C)   → *Bad*

205 | Generate the rest of backbone atoms CB, O (HN and HCA optional). Do vdw check with all other atoms after each atom is added.   → *Bad*

206 | Modeling the SS bond and record the SS coordinate pairs for later (the final) vdw check.   → *Bad*

*Continues to next page*

# FIG. 13A

16/21

*Continues from previous page*

**207** — Add rotamers (four options) to each residue except CYS. Do vdw check after each rotamer (a group of atoms) is added.

**208** — Obtain all the none-vdw-violation rotamer for each residue. Obtain all the **combinations of the rotamer**, and also do rotamer pair vdw check.

No

**209** — Write the backbone PDB file in one file, and write all the rotamer combinations into another file

**210** — Three options to add rotamers. Do vdW check for all of them    No

**211** — Do vdw check for all of the SS pairs with the complete conformer. Pick up the best SS bond.    Bad

**212** — Write the PDB file, backbone angles, all the input and statistic information file

**213** — Do "binding" test for each conformer with the receptor.

**214** — Calculate the SAS-based energy for each conformer. $\Delta G$; Probability $e^{(-\Delta G/RT)}$; Partition $\Sigma e^{(-\Delta G/RT)}$

Note:
Three options for atom pair Van der Waals contact distances check.
Option 1, normal distance; 2, extreme distance; 3, calibrated distance.
*The backbone atoms are separated from side chain atoms, when doing vdw check.*

## FIG. 13B

17/21

## A flow chart of the program MPMOD
## (the slow mod)

300 | Input: peptide sequence, disulfide bond connectivity and other parameters

301 | Randomly generate Φ, Ψ, ω in the allowed regions of the Ramachandran maps (Gly, Pro, CB-branched, and the rest of the amino acid). Assign the angles to each residue of the backbone.

302 | Generate backbone atoms N, CA, C starting from three given atoms.   Bad →

303 | Generate backbone atoms CB, O (HN and HCA optional). Do vdw check with all other atoms after each atom is added.   Bad →

304 | Add rotamers (three options) to each residue including CYS. Do vdw check after each rotamer (a group of atoms) is added. (See separate page);   Bad →

*Continues to next page*

FIG. 14A

18/21

*Continues from previous page*

305 |
1, Check the CA-CA and CB-CB distance pairs.

2, Modeling disulfide bonds record the SS pairs.

3, Do vdw check for the SS pairs with the complete conformer

No →

308 |
Record the number of conformer that can not form SS bond.

Link Corex program and calculate the SAS-based energy $\Delta G$ for each conformer.
Probability $e^{(-\Delta G/RT)}$
Partition $\Sigma e^{(-\Delta G/RT)}$

↓ Yes

306 |
Record the number of conformer SS bond.
Calculate the SAS-based energy $\Delta G$ for each conformer.
Probability $e^{(-\Delta G/RT)}$
Partition $\Sigma e^{(-\Delta G/RT)}$

307 |
Write the PDB file of each conformer (xyz-ss.pdb).
Do all the statistics (mod)

FIG. 14B

19/21

## The loop generation of program MPMOD

400 | Input the two residue numbers of the flexible loop of the protein and the accuracy.

401 | Randomly generate Φ, Ψ, ω in the allowed regions of the Ramachandran maps (Gly, Pro, CB-branched, and the rest amino acid).          *Bad*

402 | Generate mainchain atoms N, CA, C, O, HN, HCA

403 | Do Ca-Ca distance check for the N & C terminals

404 | Minimize the N and C terminal of the conformer to be the same as the cutting part of the protein, by tweaking the diheral angles.

405 | Check the handness of the conformer (It should be the same as the cutting parts of the target protein

406 | Do mainchain atom pairs vdW check.

407 | Align the conformer to the target

408 | Do mainchain and protein atom pairs vdW check.

409 | Add sidechain do vdW check for all of atom pairs.

410 | Write the closed pdb file

## FIG. 15

20/21

## Modeling disulfide bond

500 | Get the coordinates of N, Cα, Cβ of the two cysteines.

501 — The distance check
Cα....Cα<7.0Å
and
Cβ....Cβ<4.0Å

No

Yes

502 — Generated SG on the circle formed by the rotation along Cα-Cβ bond. The range of the rotation angle $\chi1$ is optional with a step at 4°
Option 1, 0 -360;
2, $|60| \pm 40°$ and $180 \pm 40°$
3, $|60| \pm 30°$ and $180 \pm 30°$

503 —

| Bond length | $S\text{-}S = 2.04 \pm 0.4Å$ |
| bond angle | $C\beta\text{-}S\text{-}S = 104 \pm 4°$ |
| | $C\alpha\text{-}C\beta\text{-}S = 114.1°$ |
| dihedral angle | $\chi^{S\text{-}S}(C\beta\text{-}S\text{-}S\text{-}C\beta) = |90| \pm 35°$ |

No

Yes

504 | SS bond is formed. Write the PDB coordinates

## FIG. 16

21/21

## The "binding" test

600

Get the following things:
1, pdb coordinates of the generated conformer and the crystal structure. 2, segment of sequence for both alignment. 3, criteria for best alignment. 4, three options for test "binding".

601

Align the conformer to the corresponding peptide crystal complex.

602

rmsd<given
and
Δ(angle)<given

No

603    Do vdW check for the conformer with the protein. (bind.f)    No

Yes

604

Calculate the SAS-based energy for each conformer. $\Delta G$; Probability $e^{(-\Delta G/RT)}$; Partition $\Sigma e^{(-\Delta G/RT)}$

605    Do statistics and print the binding information

## FIG. 17

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/09051

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(7)  :  G01N 33/48
US CL  :  702/19

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 702/19

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Please See Continuation Sheet

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| Y | US 6,188,965 B1 (MAYO et al) 13 February 2001, column 2, lines 15-39; column 7, lines 26-51; column 8, lines 15-28; column 10, lines 25-28; column 11, lines 48-62; column 14, lines 51-56. | 1-15, 48-53, 55 |
| Y | US 4,908,773 A (PANTOLIANO et al) 13 March 1990, column 3, lines 35-68 and column 4, lines 1-64; column 6, lined 8-56; column 7, lines 24-40; column 11, lines 65-67 and column 12, lines 1-16; column 15, lines 20-45. | 1-15, 48-53, 55 |
| Y | US 4,939,666 A (HARDMAN) 03 July 1990, column 3, lines 16-35; column 8, lines 24-39; column 11, lines 54-65; column 20, lines 32-51. | 1-15, 48-53, 55 |
| Y | US 5,747,654 A (PASTAN et al) 05 May 1998, column 1, lines 61-66; column 10, lines 62-66; column 13, lines 64-67 and column 14, lines 1-18; column 15, lines 36-64. | 1-15, 55 |
| Y | US 6,127,524 A (CASIPIT et al) 03 October 2000, column 1, lines 44-59; column 9, lines 61-67; column 10, lines 42-55; column19, lines 11-18; column 22, lines 49-56; column 23, lines 17-25. | 1-15, 48-53, 55 |

☒ Further documents are listed in the continuation of Box C.       ☐     See patent family annex.

| * | Special categories of cited documents: | "T" | later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention |
|---|---|---|---|
| "A" | document defining the general state of the art which is not considered to be of particular relevance | | |
| "E" | earlier application or patent published on or after the international filing date | "X" | document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone |
| "L" | document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) | "Y" | document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art |
| "O" | document referring to an oral disclosure, use, exhibition or other means | | |
| "P" | document published prior to the international filing date but later than the priority date claimed | "&" | document member of the same patent family |

| Date of the actual completion of the international search | Date of mailing of the international search report |
|---|---|
| 16 July 2002 (16.07.2002) | **2 6 AUG 2002** |
| Name and mailing address of the ISA/US<br>Commissioner of Patents and Trademarks<br>Box PCT<br>Washington, D.C. 20231<br>Facsimile No. (703)305-3230 | Authorized officer<br><br>Nikolai M Galitsky, PhD.<br><br>Telephone No. (703)308-0196 |

Form PCT/ISA/210 (second sheet) (July 1998)

## C. (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| Y | HELLINGA et al. Optimal Sequence Selection in Proteins of Known Structure by Simulated Evolution. Proceedings of the National Academy of Sciences. June 1994, Volume 91, pages 5803-5807. See entire document. | 55 |
| A | BAKER et al. Protein Structure Prediction and Structural Genomics. Science. 05 October 2001, Volume 294, pages 93-96. See entire document. | 1-15, 48-53, 55 |
| Y, P | US 6,226,603 B1 (FREIRE et al) 01 May 2001, column 12, lines 56-63; column 14, lines 48-60; column 16, lines 6-15. | 1-15, 48-53, 55 |

Form PCT/ISA/210 (second sheet) (July 1998)

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/09051

| Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet) |
| --- |

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claim Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claim Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claim Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

| Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet) |
| --- |

This International Searching Authority found multiple inventions in this international application, as follows:
Please See Continuation Sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: 1-15, 48-53 and 55

Remark on Protest     ☐ The additional search fees were accompanied by the applicant's protest.
                      ☐ No protest accompanied the payment of additional search fees.

Form PCT/ISA/210 (continuation of first sheet(1)) (July 1998)

**BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING**

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group I, claims 1-15, 48-53 and 55, drawn to a computer-assisted method for use in modifying protein with step of modelling disulfide bonds.

Group II, claims 16-22, drawn to a computer-assisted method for use in modifying protein with step of determining disulfide bonds.

Group III, claims 23 - 26. 56 and 57, drawn to a computer-assisted method for use in modifying protein that calculates linear conformation is COREX with the step of generating random conformational angeles in allowed region of Ramachandran maps.

Group IV, claims 27-38, drawn to a method for determining the rate of disulfide bond loop closure in a protein.

Group V, claims 39 - 43, and 47, drawn to a method for assessing the binding affinity of a protein to a template molecule.

Group VI, claims 44 - 46, and 54, drawn to a method for modifying protein by evaluating an X-ray crystal structure or NMR solution structure.

The inventions listed as Groups I - VI do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

The special technical features for the inventions of Group I is a computer-assisted method for use in modifying protein with step of modelling disulfide bonds.

The special technical features for the inventions of Group II is a computer-assisted method for use in modifying protein with step of determining disulfide bonds.

The special technical features for the inventions of Group III is a computer-assisted method for use in modifying protein that calculates linear conformation is COREX with the step of generating random conformational angeles in allowed region of Ramachandran maps.

The special technical features for the inventions of Group IV a method for determining the rate of disulfide bond loop closure in a protein.

The special technical features for the inventions of Group V is a method for assessing the binding affinity of a protein to a template molecule.

The special technical features for the inventions of Group VI is a method for modifying protein by evaluating an X-ray crystal structure or NMR solution structure.

**Continuation of B. FIELDS SEARCHED Item 3:**

WEST, STN, Non-Patent Literature.

Search terms: protein, modifying, modeling, design, docking, disulfide, conformer, rotamer, minituarization, entropy, backbone, Wander-Waals, COREX, minimization, computer, software, databse, program.

# This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

## BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

❑ **BLACK BORDERS**

❑ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**

☑ **FADED TEXT OR DRAWING**

❑ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**

❑ **SKEWED/SLANTED IMAGES**

❑ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**

❑ **GRAY SCALE DOCUMENTS**

❑ **LINES OR MARKS ON ORIGINAL DOCUMENT**

❑ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**

❑ **OTHER:** _____

## IMAGES ARE BEST AVAILABLE COPY.
As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

THIS PAGE BLANK (USPTO)